PROCEEDINGS

ICMPC14

JULY 5 - 9, 2016
Hyatt Regency Hotel
San Francisco, California

International Conference on Music Perception and Cognition

14th Biennial Meeting

Proceedings of the

14th International Conference on Music Perception and Cognition

Sponsors

Conference Information

Table of Contents

Author Index

Search

Proceedings of the 14th International Conference on Music Perception and Cognition. ISBN 1 876346 65 5,

SPONSORS

sempre:

Society for Education, Music and Psychology Research

UC MERCI

NSF

National Science Foundation

SMPC

Society for Music Perception and Cognition

EuroMov

Thank you for your support!

WELCOME from ICMPC14

We are pleased you could join us for the 14th International Conference on Music Perception and Cognition (ICMPC14) and we welcome you to San Francisco. It has been a distinct privilege for us to host this event and we look forward to your involvement. We believe the historical success of ICMPC stems from the rich diversity of the attendees, who despite differences in geographical location, educational background, and professional focus, come together biennially as a commitment to advance this truly interdisciplinary field of music perception and cognition. This year, ICMPC14 will bring together over 500 attendees from 32 different countries who boast a myriad of specializations, including psychologists, educators, theorists, therapists, performers, computer scientists, ethnomusicologists and neuroscientists. And what better place to celebrate this diversity than San Francisco – a city renowned for its melding of cultures and ideas. As such, ICMPC14 is poised to continue the great tradition of blending worldc class interdisciplinary research presentations with stimulating social opportunities, both inside and outside the conference venue. On behalf of the organizing committee, the conference advisory board, and the conference staff and volunteers, we hope you enjoy your time at ICMPC14 and that you feel at home here in San Francisco. Please do not hesitate to ask if there is anything we can do enhance your conference experience.

Sincerely,

Theodore Zanto
ICMPC14 Conference Chair
University of California San Francisco

WELCOME from SMPC

Dear ICMPC 2016 Delegates,

I am happy to welcome you to the 14th biennial International Conference on Music Perception and Cognition, hosted this year in North America and sponsored in part by the Society for Music Perception and Cognition. SMPC is an organization of scientists and scholars conducting research in music perception and cognition, broadly construed. Our membership, which numbers over 500, represents a broad range of disciplines, including music theory, psychology, psychophysics, linguistics, neurology, neurophysiology, ethology, ethnomusicology, artificial intelligence, computer technology, physics and engineering, and we are witnessing an unprecedented proliferation of research labs and graduate programs dedicated to the scientific study of music. We are historically a North American society, however we have members around the globe, including Europe, Asia and Australia, and we are proud to participate in this year's meeting. To learn more about SMPC, please visit our website, www.musicperception.org, where you can read recent news, learn about upcoming events, find information about our laboratories and graduate programs, and watch videos and podcasts that feature the latest research in music perception and cognition.

I wish to express my sincere thanks to Dr. Theodore Zanto and all the organizers for planning and coordinating this year's meeting in San Francisco. Conferences like ICMPC are essential to the dissemination and coordination of research in our growing field, and Dr. Zanto has done a spectacular job of putting together this meeting in one of the most beautiful cities in North America. I am also pleased to announce that SMPC's next biennial meeting will be hosted at University of California San Diego, with Dr. Sarah Creel as conference chair. The tentative dates for the meeting are Jul 31 – Aug 4, and we hope to have these and other details finalized soon.

Best wishes for an exciting and productive meeting,

Ed Large
President
Society for Music Perception and Cognition

CONFERENCE TEAM

Organizing Committee

Theodore Zanto, Conference Chair
University of California San Francisco

Takako Fujioka
Stanford University

Petr Janata
University of California Davis

Julene Johnson
University of California San Francisco

Jonathan Berger
Stanford University

Jessica Slater
Northwestern University

Chris Chafe
Stanford University

Program Coordinators

Vinith Johnson
University of California San Francisco

Di Zhao
University of California San Francisco

Harrison Conrad
University of California San Francisco

Event Planners

Tara Miller
TME Events, Inc.

Kerry Bosch
TME Events, Inc.

Proceedings Coordinator

George Vokalek
Causal Productions

Volunteers

Angela Nazarian

Ariadna Saenz

Ben Kubit

Camille Nawawi

Daniel Myles

Jade Verheijen

Marijn Coers

Namita Padgaonkar

Tessa Overboom

Wan-Yu Hsu

CONFERENCE COMMITTEES

ICMPC Executive Committee

Steven Demorest (Chair, ICMPC11)

Emilios Cambouropoulos (ICMPC12)

Moo Kyoung Song (ICMPC13)

Nikki Rickard (APSCOM)

Jane Ginsborg (ESCOM)

Edward Large (SMPC)

ICMPC Advisory Board

Sun-Hee Chang (President APSCOM)

Jane Ginsborg (President ESCOM)

Edward Large (President SMPC)

Diana Deutsch (Co-Organizer 1st ICMPC)

Roger Kendall (Organizer 2nd ICMPC)

Irene Deliege (Organizer 3rd ICMPC)

Eugenia Costa-Giomi (Organizer 4th ICMPC)

Bruce Pennycook (Co-Organizer 4th ICMPC)

Suk Won Yi (Organizer 5th ICMPC)

John Sloboda (Organizer 6th ICMPC)

Susan O'Neill (Co-Organizer 6th ICMPC)

Kate Stevens (Organizer 7th ICMPC)

Scott Lipscomb (Organizer 8th ICMPC)

Anna Rita Addessi (Co-Organizer 9th ICMPC)

Mayumi Adachi (Organizer 10th ICMPC)

Steven Demorest (Organizer 11th ICMPC)

Emilios Cambouropoulos (Organizer 12th ICMPC)

Moo Kyoung Song (Organizer 13th ICMPC)

ICMPC14 / SEMPRE Young Investigator Award Committee

Elaine Chew (Queen Mary University London)

Steve Demorest (Northwestern University)

Jay Dowling (University of Texas, Dallas)

Takayuki Nakata (Future University, Hakodate)

Scientific Advisory Board

Alexandra Lamont (Keele University)

Amy Graziano (Chapman University)

Andrea Halpern (Bucknell University)

Anna Rita Addessi (University of Bologna)

Annabel Cohen (University of Prince Edward Island)

Barbara Tillmann (Lyon Neuroscience Research Center)

Charles Limb (University of California San Francisco)

David Temperley (University of Rochester)

Devin McAuley (Michigan State University)

Elizabeth Margulis (University of Arkansas)

Elvira Brattico (Aarhus University)

Emilios Cambouropoulos (Aristotle University of Thessaloniki)

Erin Hannon (University of Nevada Las Vegas)

Erkki Huovinen (University of Minnesota)

Eugenia Costa-Giomi (Ohio State University)

Glenn Schellenberg (University of Toronto at Mississauga)

Henkjan Honing (University of Amsterdam)

Jane Ginsborg (Royal Northern College of Music)

Jessica Grahn (Western University)

Joel Snyder (University of Nevada Las Vegas)

John Iversen(University of California San Diego)

John Sloboda (Guildhall School of Music & Drama)

Jonathan Berger (Stanford University)

Jukka Louhivuori (University of Jyväskylä)

Julene Johnson (University of California San Francisco)

Justin London (Carleton College)

Kate Stevens (University of Western Sydney)

Kyung Myun Lee (Seoul National University)

Laurel Trainor (McMaster University)

Lauren Stewart (Goldsmith, University of London)

Lola Cuddy (Queen's University)

Marc Leman (University of Ghent)

Marcus Pearce (Queen Mary, University of London)

Mark Schmuckler (University of Toronto Scarborough)

Morwaread Farbood (New York University)

Nikki Rickard (Monash University)

Peter Keller (University of Western Sydney)

Peter Pfordresher (University at Buffalo State University of New York)

Peter Vuust (Aarhus University)

Petr Janata (University of California Davis)

Petri Toiviainen (University of Jyväskylä)

Psyche Loui (Wesleyan University)

Ramesh Balasubramaniam (University of California Merced)

Reinhard Kopiez (Hanover University of Music)

Reyna Gordon (Vanderbilt University)

Richard Ashley (Northwestern University)

Rita Aiello (New York University)

Robert Gjerdingen (Northwestern University)

Sarah Creel (University of California San Diego)

Scott Lipscomb (University of Minnesota)

Stefan Koelsch (Freie Universität Berlin)

Stephen McAdams (McGill University)

Steven Brown (McMaster University)

Steven Morrison (University of Washington)

Suk Won Yi (Seoul National University)

Takako Fujioka (Stanford University)

Theodore Zanto (University of California San Francisco)

Tom Collins (de Montfort University)

Werner Goebl (University of Music and Performing Arts)

William Forde Thompson (Macquarie University

AWARDS

ICMPC14 / SEMPRE Young Investigator Award

This award is granted to research submissions of exceptional quality that were undertaken by either graduate students or new researchers in the field. Each applicant for the award submitted a paper of up to 3000 words that was reviewed by the Young Investigator Award selection committee. The award winner received $1500 in travel and housing support and complimentary registration courtesy of SEMPRE & ICMPC14. The honorable mention recipient received complimentary registration and $1000 toward travel expenses.

Young Investigator Award Winner

Pauline Larrouy-Maestri (Max-Planck Institute)

Perception of pitch accuracy in melodies: A categorical or continuous phenomenon?

Pitch & Tonal Perception 1: Wednesday, 9:00 am, Marina

Honorable Mention

Tan-Chyuan Chin (The University of Melbourne)

Predicting emotional well-being: The roles of emotional sensitivity to music and emotion regulation using music

Music & Health / Well-Being 2: Thursday, 9:30 am, Marina

NSF Travel Awards

Four travel awards were presented to local (U.S.) graduate students who submitted an exceptional research paper of up to 3000 words. Awardees received complete travel, housing and registration support, courtesy of the National Science Foundation. Congratulations to the NSF Travel Award winners:

Isabelle Arseneau-Bruneau (University of Texas at Austin)

Focus Of Attention And Learning During Trumpet Musical Performance Music

Music Education 4: Saturday, 4:00 pm, Seacliff A-C, Sa29

Samuel Mehr (Harvard University)

For 5-Month-Old Infants, Melodies Are Social

Social Psychology of Music 2: Thursday, 10:00 am, Seacliff D

Brooke Okada (University of Maryland)

Musical Experience and Executive Function

Memory & Music 3: Saturday, 11:45 am, Bayview B

Kelly Whiteford (University of Minnesota)

Robust training effects in congenital amusia

Pitch & Tonal Perception 3: Thursday, 4:00 pm, Seacliff A-C, Th50

SEMPRE Travel Awards

ICMPC14 would like to thank the Society for Education, Music, and Psychology Research (SEMPRE) for their unprecedented generosity in supporting ICMPC attendees. This year, SEMPRE provided over $23,000 in awards to conference attendees! Awardees received various amounts based on need. Congratulations to the 50 SEMPRE Travel Award winners:

Stefanie Acevedo
Lauren Amick
Valentin Begel
Claire Castle
Nisha Chandrasekaran
Laura Cirelli
Shawn Condon
Esther Coorevits
Jose Davison da Silva Jr
Maria Fasano
Elisa Fromboluti
Sarah Gates
Rebecca Gelding
Steffen Herff
Yu-Fen Huang
Patricia Izbicki
Mendel Kaelen
Anna Kasdan
Savvas Kazazis
Haley Kragness
Adam Linson
Xuejing Lu
Fiona Manning
Manuela Marin
Avantika Mathur
Eva Matlschweiger
Siddharth Mehrotra
Karli Nave
Jason Noble
Tessa Overboom
Ana Pereira
Jasmin Pfeifer
Rita Pfeiffer
Olivia Podolak
Tamar Regev
Dawn Rose
Dave Rosen
Suzi Ross
Giovanni Sala
Pathmanesan Sanmugeswaran
Sabrina Sattmann
Kirsten Smayda
Christopher (Kit) Soden
Swathi Swaminathan
Cecilia Taher
Karen Thomas
Katherine Thompson
Michael Weiss
Xin (Olivia) Wen
Courtney Williams

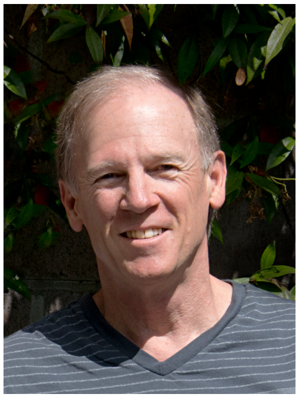
Thank you to the sponsors of the ICMPC14 Awards:

sempre:
Society for Education, Music and Psychology Research

ICMPC14

NSF National Science Foundation

OPENING KEYNOTE



Bach to the blues: Music, color, emotion and synesthesia

Bayview A+B, Tuesday at 9 am

Steve Palmer, Ph.D.
Professor, Department of Psychology
University of California Berkeley

I will discuss cross-modal associations from music to colors for several kinds of music, including classical orchestral pieces by Bach, Mozart, and Brahms, single-line piano melodies by Mozart, and 34 different genres of popular music from jazz to heavy-metal to salsa to country-western. When choosing the 3 colors (from 37) that "went best" with selections of classical music, choices by 48 US non-synesthetes were highly systematic: e.g., faster music in the major mode was strongly associated with more saturated, lighter, yellower colors. Further results strongly suggest that emotion mediates these music-to-color associations: e.g., the happy/sad ratings of the classical orchestral selections were highly correlated .97 with the happy/sad ratings of the colors chosen as going best with them (Palmer, Schloss, Xu, & Prado-León, PNAS, 2013). Cross-cultural results from Mexican participants for the same classical selections were virtually identical to those from US participants. Analyses of the 34 genres data further reveals that color choices can be better predicted by 3 emotional dimensions (happy/sad, agitated/calm, and like/dislike) than by 8 perceived musical dimensions (electronic/acoustic, fast-strong/slow-weak, harmonious/disharmonious, etc.) Strong emotional effects were also present for two-note musical intervals, and weaker emotional effects for the timbre (or tone color) of individual instruments. Similar experiments were conducted with 12 music-to-color synesthetes, except that they chose the 3 colors (from the same 37) that were most similar to the colors they actually experienced while listening to the same musical selections. Synesthetes showed clear evidence of emotional effects for some musical variables (e.g., major versus minor) but not for others (e.g., slow versus fast tempi in the piano melodies). The complex nature of similarities and differences between synesthetes' color experiences and non-synesthetes' color associations will be discussed.

CLOSING KEYNOTE

Physically embodied music cognition

Bayview A+B, Saturday at 5 pm

Edward Large, Ph.D.
Professor, Department of Psychology
University of Connecticut

Music involves physical interactions with musical instruments, acoustic transmission of sound waves, vibration of peripheral auditory structures and dynamic coordination of whole body movements. At the same time, music engages high-level cognitive and affective processes, and is often compared to language in terms of its structural and cognitive complexity. How can we bridge the gap between the raw physical experience of music and the complex cognitive computations that are assumed to enable it? In this talk, I propose a radically embodied approach to music cognition, consistent with the enactive approach of Varela and the ecological psychology of Gibson. The “radical” part is that it appeals to physical principles to explain music at all levels of description: patterns of activity in the brain, patterns of action in the body, and patterns of interaction among groups of individuals. At the neural level, emergent dynamic patterns — as opposed to special purpose circuits — are considered to be the fundamental elements, and resonance — rather than computation — is considered to be the fundamental operation. However, this approach is not neurocentric. It recognizes that music cuts across brain–body–world divisions and it provides the concepts and tools necessary to bridge this divide. I begin with rhythm, reviewing theoretical and empirical results that show why rhythm perception-action is best understood as resonance of the brain and body to sound. Using the example of synchronizing metronomes, I explain how we can use a single language — nonlinear dynamical systems — to cut across brain-body-world boundaries. I show how this approach replaces abstract cognitive “computation” with resonance of emergent physical patterns. Finally, I describe how the radically embodied approach leads to a view of music as a natural, emergent phenomenon of the physical world.

Table of Contents

Aesthetic Perception & Response 1

Wednesday, 10:30–12:30, Marina, Session Chair: Bill Thompson

Aesthetic Perception & Response 2

Friday, 08:30–10:30, Marina, Session Chair: Andrea Halpern

Aesthetic Perception & Response 3

Saturday, 14:00–15:00, Marina, Session Chair: Erkki Huovinen

Aesthetic Perception & Response 4 (Posters)

Tuesday , 16:30, Seacliff A–C

Cognitive Modeling of Music 1

Tuesday, 15:30–16:15, Seacliff D, Session Chair: Joshua Albrecht

Cognitive Modeling of Music 2

Friday, 11:00–12:20, Bayview A

Symposium: Beneath (or Beyond) the Surface: Corpus Studies of Tonal Harmony
Discussants: David Sears (McGill University); Claire Arthur (Ohio State University)

Cognitive Musicology 2

Saturday, 08:30–10:30, Seacliff D, Session Chair: Richard Ashley

---

Cognitive Musicology 3 (Posters)

Thursday, 16:00, Seacliff A–C

Composition & Improvisation 1

Tuesday, 14:00–15:30, Bayview A, Session Chair: Andrew Goldman

Composition & Improvisation 2 (Posters)

Tuesday, 16:30, Seacliff A–C

Cross-Cultural Studies of Music 1

Saturday, 15:00–15:30, Seacliff D, Session Chair: Steven Demorest

Cross-Cultural Studies of Music 2 (Posters)

Saturday, 16:00, Seacliff A–C

History of Music Cognition 1

Friday, 11:00–12:30, Seacliff D

Symposium: Empirical Musicology, 10 Years On: New Ways to Publish and the Empirical Paradigm in Music Research
Discussant: Daniel Shanahan (Louisiana State University)

History of Music Cognition 2

Saturday, 11:00–12:30, Seacliff D

Symposium: Perspectives on the History of Music Cognition
Discussants: Amy Graziano (Chapman University); Julene Johnson (University of California San Francisco)

Memory & Music 1

Tuesday, 14:00–15:30, Marina, Session Chair: Sandra Trehub

Memory & Music 2

Tuesday, 15:30–16:30, Marina, Session Chair: Roger Dean

Memory & Music 3

Saturday, 11:00-12:00, Bayview B, Session Chair: Daniel Müllensiefen

Memory & Music 4

Saturday, 15:00–16:00, Marina, Session Chair: L. Robert Slevc

Memory & Music 5 (Posters)

Wednesday, 14:00, Seacliff A–C

Music & Aging (Posters)

Saturday, 16:00, Seacliff A–C

Music & Emotions 1

Tuesday, 10:30–11:30, Bayview B, Session Chair: E. Glenn Schellenberg

Music & Emotions 2

Thursday, 08:30–10:30, Bayview B, Session Chair: Fernando Bravo

Music & Emotions 3

Saturday, 14:00–15:00, Bayview B, Session Chair: Michael Schutz

---

Music & Emotions 4

Saturday, 15:00–16:00, Bayview B, Session Chair: Tan-Chyuan Chin

---

Music & Emotions 5 (Posters)

Tuesday, 16:30, Seacliff A–C

Music & Emotions 6 (Posters)

Wednesday, 14:00, Seacliff A–C

Music & Emotions 7 (Posters)

Saturday, 16:00, Seacliff A–C

Music & Evolution (Posters)

Tuesday, 16:30, Seacliff A–C

Music & Health / Well-Being 1

Wednesday, 08:30–09:50, Bayview A

Symposium: Choral Singing and Well-Being
Discussant: Jacques Launay (Brunel University of London)
Convener: Annabel Cohen (University of Prince Edward Is)

Music & Health / Well-Being 2

Thursday, 08:30–10:30, Marina, Session Chair: Nikki Rickard

Music & Language 3

Wednesday, 08:30–09:50, Bayview B

Symposium: Musical Rhythm in Typical and Atypical Language Development
Discussant: Sarah Creel (University of California San Diego)
Conveners: J. Devin McAuley (Michigan State University); Reyna Gordon (Vanderbilt University)

Music & Language 4

Thursday, 08:30–10:30, Bayview A, Session Chair: Reyna Gordon

Music & Language 5

Saturday, 14:00–14:45, Bayview A, Session Chair: Psyche Loui

Music & Language 6

Saturday, 15:00–15:45, Bayview A, Session Chair: Ao Chen

Music & Language 7 (Posters)

Wednesday, 14:00, Seacliff A–C

Music & Language 8 (Posters)

Saturday, 16:00, Seacliff A–C

Music & Meaning 1

Thursday, 14:00–15:00, Bayview B, Session Chair: Martin Norgaard

Music & Movement 1

Tuesday, 11:30–12:30, Bayview A, Session Chair: Petri Toiviainen

Music & Movement 2

Thursday, 11:00–12:30, Seacliff D, Session Chair: Marc Thompson

Music & Movement 3

Friday, 08:30–10:00, Bayview A, Session Chair: Roger Chaffin

Music & Movement 4

Saturday, 11:00–12:30, Bayview A

Symposium: Embodiment in Electronic Dance Music
Discussants: Hauke Egermann (University of York); Marc Thompson (University of Jyväskylä)

Music & Movement 5 (Posters)

Wednesday, 14:00, Seacliff A–C

Music & Movement 6 (Posters)

Thursday, 16:00, Seacliff A–C

Music & Neuroscience 1

Tuesday, 15:30–16:30, Bayview B, Session Chair: John Iversen

Music & Neuroscience 2

Wednesday, 10:30–12:00, Bayview B, Session Chair: Bernhard Ross

Music & Neuroscience 3

Friday, 11:00–12:30, Bayview B

Symposium: Exploring Music & Language Processing and Emotions with ECoG
Discussant: Petr Janata (UC Davis)
Conveners: Christian Mikutta (University of California Berkeley); Robert Knight (University of California Berkeley)

Music & Neuroscience 7 (Posters)

Thursday, 16:00, Seacliff A–C

Music & Personality 1

Tuesday, 14:00–15:30, Bayview B, Session Chair: Birgitta Burger

Music & Personality 2 (Posters)

Saturday, 16:00, Seacliff A–C

Music Education 1

Tuesday, 11:30–12:30, Marina, Session Chair: Reinhard Kopiez

Music Performance 2

Wednesday, 16:30–17:30, Marina, Session Chair: Stacey Davis

Music Performance 3

Thursday, 14:00–15:30, Seacliff D, Session Chair: Peter Keller

Music Performance 4

Friday, 08:30–10:00, Seacliff D, Session Chair: Kyung Myun Lee

Music Performance 5

Saturday, 14:00–15:00, Seacliff D, Session Chair: Neta Spiro

Music Performance 6 (Posters)

Wednesday, 14:00, Seacliff A–C

Music Therapy 1

Tuesday, 11:30–12:30, Seacliff D, Session Chair: Charles Limb

Music Therapy 2 (Posters)

Thursday, 16:00, Seacliff A–C

Music, Health, Science & Industry 1

Wednesday, 10:30–12:10, Bayview A

Music, Health, Science & Industry 2

Wednesday, 15:00, Bayview A

Panel Discussion: The Intersection of Art, Medicine, Science and Industry: Improving Public Health and Well-Being Through Music and Technology

Discussant: Benoit Bardy (University of Montpellier)

Convener: Theodore Zanto (University of California San Francisco)

Panel: Adam Gazzaley (University of California San Francisco)
Michael Winger (The Recording Academy, SF)
Matt Logan (UCSF Benioff Children's Hospital)
Ketki Karanam (The Sync Project)

Music, Health, Science & Industry 3

Wednesday, 16:30–17:30, Bayview A

Symposium: Leveraging Large-Scale Industrial Data to Study Musical Engagement
Discussant: Douglas Eck (Google)
Convener: Blair Kaneshiro (Stanford University)

Musical Development 1

Tuesday, 10:30–11:30, Marina, Session Chair: Adam Tierney

Musical Timbre 1

Friday, 08:30–10:30, Bayview B, Session Chair: Jason Noble

Pitch & Tonal Perception 1

Wednesday, 08:30–10:30, Marina, Session Chair: Carol Krumhansl

Pitch & Tonal Perception 2

Wednesday, 15:00–16:00, Seacliff D, Session Chair: Pauline Larrouy-Maestri

Pitch & Tonal Perception 3

Thursday, 11:00–12:30, Bayview B

Pitch & Tonal Perception 4

Saturday, 08:30–10:30, Bayview A, Session Chair: Marianne Ploger

Pitch & Tonal Perception 5 (Posters)

Thursday, 16:00, Seacliff A–C

Pitch & Tonal Perception 6 (Posters)

Saturday, 16:00, Seacliff A–C

Rhythm, Meter & Timing 1

Tuesday, 11:30–12:30, Bayview B, Session Chair: Eric Clarke

Rhythm, Meter & Timing 2

Wednesday, 15:00–16:00, Bayview B, Session Chair: Simone Dalla-Bella

Rhythm, Meter & Timing 3

Wednesday, 16:30–18:00, Bayview B, Session Chair: Erin Hannon

Rhythm, Meter & Timing 4

Thursday, 14:00–15:30, Bayview A, Session Chair: Justin London

Rhythm, Meter & Timing 5 (Posters)

Tuesday, 16:30, Seacliff A–C

Rhythm, Meter & Timing 6 (Posters)

Thursday, 16:00, Seacliff A–C

Rhythm, Meter & Timing 7 (Posters)

Saturday, 16:00, Seacliff A–C

Social Psychology of Music 1

Wednesday, 10:30–12:30, Seacliff D, Session Chair: Jacques Launay

Social Psychology of Music 2

Thursday, 08:30–10:30, Seacliff D, Session Chair: Jonna Vuoskoski

Social Psychology of Music 3 (Posters)

Wednesday, 14:00, Seacliff A–C

Social Psychology of Music 4 (Posters)

Saturday, 16:00, Seacliff A–C

Testing the Robustness of the Timbre Toolbox and the MIRtoolbox

Savvas Kazazis[1], Nicholas Esterer[1], Philippe Depalle[1] and Stephen McAdams[1]
[1]Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT), Schulich School of Music, McGill University, Montreal, Canada

The Timbre Toolbox and MIRtoolbox are the most popular Matlab toolboxes that are used for audio feature extraction within the MIR community. They have been recently evaluated according to the number of presented features, the user interface and computational efficiency, but there have rarely been performance evaluations of the accuracy of the extracted features. In this study we evaluate the robustness of audio descriptors these toolboxes have in common and that are putatively related to timbre. For this purpose we synthesized various soundsets using additive synthesis, which allows for a direct computation of the audio descriptors. We evaluate toolbox performance by analyzing the soundsets with each toolbox and calculating the normalized RMS error between their output and the theoretical values. Each soundset was designed to exhibit specific sound qualities that are directly related to the descriptors being tested. In this way we are able to systematically test the performance of the toolboxes by tracking under which circumstances certain audio descriptors are poorly calculated. For the analysis we used the default settings of each toolbox and the summary statistics from the frame-by-frame analysis were derived using the median values. For spectral descriptors the Timbre Toolbox performs better and on some soundsets outperforms the MIRtoolbox with the short-term Fourier transform power representation being the most robust. The Timbre Toolbox failed to analyze some sounds using the harmonic representation even though all sounds were strictly harmonic. The MIRtoolbox's calculations of spectral centroid on some sounds and spectral irregularity on a specific soundset proved to be numerically unstable. For descriptors that are based on the estimation of the amplitude envelope both toolboxes perform almost equally but poorly. We noticed that in this case the errors depend both on the length of the attack or decay times and on the shape of the slopes.

Musicians do not experience the same impairments as non-musicians during postural manipulations in tactile temporal order judgment tasks

James Brown[1], Emily Orchard-Mills[1], and David Alais[1]
[1]School of Psychology, Faculty of Science, University of Sydney, Australia

In a temporal order judgment (TOJ) task, participants are required to judge the order of two or more stimuli presented close together in time (e.g. which stimulus occurred first). Studies concerning tactile perception will often use TOJ tasks to make inference on the temporal and spatial limits of the tactile sense (i.e. how sensitive or precise the tactile sense is at locating a stimulus in time or external space). When a tactile stimulus is received at the skin, it is thought to be processed with regard to two properties: somatotopic (i.e. what part of the skin received the stimulus) and external spatial (i.e. where that part of the skin was in external space when it received the stimulus). Postural manipulations (e.g. crossing the hands into opposite external hemifields) can cause a conflict between these two properties and impair the temporal and spatial limits of the tactile sense in that context. Given their purpose, TOJ tasks can therefore help assess the impairment. Previous studies by other researchers, and pilot studies by the current authors, suggest that musicians exhibit finer temporal and spatial abilities when it comes to the tactile sense – it is therefore of interest to the current researchers if the temporal and spatial impairments associated with postural manipulations will affect musicians the same as non-musicians. A TOJ task with both normal and postural manipulation conditions was conducted on a group of musicians and non-musicians. Musicians performed significantly better than non-musicians in all conditions, with the largest differences in conditions with conflict between somatotopic and external spatial properties. Possible reasons for the differences in performance, and the apparent resilience of musicians to impairment by postural manipulation, are briefly discussed.

Semantic Dimensions of Sound Mass Fusion

Stephen McAdams[1], Jason Noble[1]
[1]Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT), Schulich School of Music, McGill University, Montreal, QC, CANADA

Sound mass, the grouping of multiple sound events or sources into perceptual units that retain an impression of multiplicity, is essential to daily auditory experience (e.g. when we recognize a crowd, traffic, or rain as a "single" sound). It is also highly relevant to musical experience: composers including Ligeti, Xenakis, and many others have exploited it at length. Such composers often describe their aesthetic aims in vividly evocative terms – clouds, galaxies, crystal lattices – but it remains unclear to what extent, if any, such metaphors relate to the experiences of listeners.

Following three experiments that examined psychoacoustic conditions predicting sound mass perception in listeners (Douglas et al, 2016), a fourth experiment examining semantic dimensions of sound mass perception is currently underway. Subjects hear 40 samples of contemporary music that may or may not exhibit sound mass fusion and rate them on a battery of scales. For each sample, participants rate: (1) its degree of sound mass fusion, (2) the degree to which it exhibits properties related to sound mass fusion (e.g. density, complexity, homogeneity), (3) the degree to which it evokes adjectival metaphors for sound mass used by scholars (e.g. volatility, stasis, formlessness), and (4) the degree to which it evokes nominal metaphors for sound mass used by composers (e.g. clouds, galaxies, crystal lattices). The results will be compared with musical / sonic analysis of the stimuli. Data analysis will reveal: (1) which psychoacoustic properties correlate best with sound mass perception, (2) to what extent listeners agree on the metaphors evoked by these samples, (3) whether or not the metaphors identified by listeners are the same ones used by scholars and composers, and (4) whether or not those metaphors correlate with identifiable properties of the stimuli. Ultimately this research will help explain the perceptual significance of sound mass fusion in contemporary music.

Aging Adults and Gender Differences in Musical Nuance Perception

Jennifer A. Bugos[1]
[1]School of Music, Center for Music Education Research, University of South Florida, Tampa, FL, USA

Background
Musical nuance, fine grain deviations in pitch, loudness, time, and tone quality in a musical phrase, is critical to listener perceptions. Research suggests that music training may influence nuance perception in young adults; however, little is known about the extent to which aging or other demographic variables such as gender, influence musical nuance perception (Bugos, Heller, & Batcheller, 2015).

Aims
The aims were to evaluate the role of gender on musical nuance perception in older adults and to compare the performance of older adults with younger adults.

Methods
One-hundred twenty healthy older adults (47 males, 73 females) were recruited from an independent living facility and screened for cognitive deficits and hearing loss. Younger adults (*n*=120) were recruited from a local urban university. Participants completed a short demographic survey and the Musical Nuance Task (MNT). The MNT is a highly reliable, 30-item measure that includes 15 items of three short musical phrases performed on the same instrument (cello, clarinet, or piano) and 15 items of three short musical phrases performed on each of these same three instruments. Two of the three performed phrases were considered the "same" in nuance and one was considered "different" in nuance.

Results
Results of a MANOVA on the MNT show significantly faster reaction times for older adult females compared to males. While no differences were found for the total number of errors on the MNT, less errors were committed by older adult females on the last 15 trials, items with multiple timbres presented. No differences were found in errors rates between young adults and older adults. However, younger adults show faster reaction times compared to older adults.

Conclusions
Nuance perception is more difficult with contrasting timbres. Gender differences in auditory processing may influence nuance perception.

Insights into the complexities of music listening for hearing aid users

Alinka Greasley[1], Harriet Crook[2], Robert Fulford[1], Jackie Salter[1]
[1]School of Music, University of Leeds, Leeds, UK
[2]Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK

Music listening has significant health and well-being benefits, including for people with all levels of hearing impairment. Digital hearing aids (HA) are optimised for speech amplification and can present difficulties for music perception. A UK-based project is currently exploring how music listening behaviour is shaped by hearing impairment and the use of HA technology. Findings from two studies will be reported. First, a clinical questionnaire explored the extent of music listening issues and the frequency and success of discussions with audiologists about music. Data from 176 HA users (age range 21-93, average=60.35, SD=18.07) showed that problems with music listening were often experienced and almost half reported that this negatively affects their quality of life. Participants described issues listening to live music performances, hearing words in songs, the loss of music from their lives and associated social isolation. Most had never talked with their audiologist about music and, for those that had, improvements were rarely reported. A second study explored listening experiences of HA users in greater depth, with the collection of pure tone audiometry to facilitate interpretation of accounts. The sample (n=22, age range 24-82, average=62.05, SD=24.07) included participants with a range of hearing impairments and levels of musical training. Participants described a variety of listening behaviours and preferences combined with different patterns of HA and Assistive Listening Device use. Analyses showed interactions between contextual factors such the individual nature of hearing loss (e.g. type, level, duration); levels of musical engagement (e.g. daily exposure, training) and contexts (e.g. recorded music at home/travelling, live performances). These studies provide new data on a poorly understood topic that will be used to inform the design of a national survey which seeks to identify patterns in the listening behaviour of a wider population of HA users.

The Perception of Auditory Distortion Products from Orchestral Crotales

Alex Chechile[1]
[1]CCRMA, Stanford University, Stanford, California, United States

On the Sensations of Tone is a series of compositions that use specific acoustic material to create a separate stream of music generated by the body. Upon the simultaneous presentation of two closely related pure-tone frequencies, auditory distortion products are produced on the basilar membrane in the cochlea and result in the perception of additional tones not present in the acoustic space. From 18th century concert music to contemporary electronic music, composers have implemented "combination tones" in pieces using instruments ranging from the violin to computer-based synthesizers. The aim of the present work was to examine the acoustic properties of the orchestral crotales within the context of combination tones and the composition *On the Sensations of Tone VIII.* The crotales are small cymbals chromatically tuned in the range of C6 to C8 (or F8), and are usually played with mallets or sometimes bowed. Pre-compositional methods involved conducting a spectral analysis in MATLAB of the upper octave crotales, calculating the appropriate distortion products, and matching the acoustic tones with synthesized frequencies. The results indicated clearly perceivable additional tones, which informed the compositional process. A case study of the piece illustrates how distortion products and virtual pitches allow for additional harmonic material. Distortion products also allow for expanded spatial depth between acoustic material and material generated in the ear. Macroscopic and microscopic listening is possible by shifting attention between the different streams, and contributes to a unique listening experience.

Vibrotactile perception of music

Frank A. Russo[1]
[1]Department of Psychology, Ryerson University, Toronto, Ontario, Canada

While music is primarily experienced in the auditory domain, many aspects of music can also be experienced through vibrotactile stimulation alone. This paper provides an overview of research in our lab that has assessed the perception of vibrotactile music. Specific features of music considered include pitch, rhythm, and timbre. We also consider differences in vibrotactile and auditory perception with regard to higher-level constructs such as melody and emotion. Finally, we consider methodological issues in this research and the manner in which the utilization of cues may change as a function of hearing loss, experience, and magnitude of stimulation.

Perceptual Learning of Abstract Musical Patterns: Recognizing Composer Style

Carolyn A. Bufford,[*1] Khanh-Phuong Thai,[2] Joselyn Ho[*], Cindy Xiong[*], Carly A. Hines[*], Philip J. Kellman[*]

[*]*Department of Psychology, University of California, Los Angeles, USA*

[1]cabufford@g.ucla.edu

ABSTRACT

How can we improve abstract pattern recognition in music? Can principles enhancing visual learning be extended to auditory stimuli, such as music? Perceptual learning (PL), improvements in the pickup of information from experience, is well-established in both vision and audition. Most work has focused on low-level discriminations, rather than extraction of relations in complex domains. In vision, recent research suggests that PL can be systematically accelerated in complex domains, including aviation, medicine, and mathematics, through perceptual and adaptive learning modules (PLMs). PLMs use brief interactive classification trials along with adaptive spacing to accelerate perceptual category learning. We developed a Composer PLM to apply these techniques to recognizing composers' styles, a challenging, abstract, PL task involving complex relations over time. We investigated whether (1) participants could learn composers' styles, and (2) whether the PLM framework could be successfully extended to rich auditory domains. On each PLM trial, participants listened to a 15-second clip of solo piano music by one of 4 composers (2 Baroque: Bach, Handel; 2 Romantic: Chopin, Schumann), attempted to identify the composer, and received feedback. Before and after multiple PLM sessions, we tested participants' ability to distinguish clips from composers in the PLM, unlearned composers in trained periods (Baroque: Scarlatti; Romantic: Mendelssohn) and composers in untrained periods (Renaissance: Byrd; post-Romantic: Debussy). At pretest, participants' sensitivity was at chance, but they showed robustly improved classification at posttest on both composer styles ($p < .01$) and period styles ($p < .001$). Results indicate that PLM training can improve participants' recognition of composers' styles, demonstrating that (1) composer style can be learned, and that (2) PL-based interventions are effective in complex auditory domains.

I. INTRODUCTION

Humans have impressive abilities to recognize structure and patterns in the world, in multiple modalities and many domains. For example, we can recognize familiar voices on the phone and identify objects at unusual sizes and angles. Music, in particular, is a useful domain in which to study pattern recognition because of its many patterns at many levels of complexity and abstraction which allow research into auditory perception across the spectrum from basic processing to high-level real-world tasks. The simplest patterns in music are specific note durations and pitches. Themes are more complex patterns, but even untrained listeners can recognize a theme at its restatement in a piece they have not heard before (Java, Kaminska, & Gardiner, 1995). The styles of composers are perhaps the most abstract and complex patterns that humans can detect in music, because these patterns transcend individual pieces or works. Research on perceptual learning as well as research specifically on composer style contribute insights into our ability to recognize composer styles, which are abstract auditory patterns.

A. Perceptual Learning

Perceptual learning (PL) refers to improvements in the pickup of information as a function of experience (Gibson, 1969). In particular, PL has been shown to accelerate the development of expert pattern recognition, contributing to the development of expertise. PL is well-established in both vision (Sagi, 2010) and audition (Sabin, Eddins, & Wright, 2012). Much work in both modalities has focused on expertise in low-level discriminations. For example, many studies have demonstrated PL in studies of pitch duration (e.g. Karmarkar & Buonomano, 2003) or orientated visual gratings (e.g. Song, Peng, Lu, Liu, & Li, 2007). However, perceptual learning also supports the extraction of relations in complex domains. In vision, recent research suggests that PL can be systematically accelerated in aviation (Kellman & Kaiser, 1994), medicine (Thai, Krasne, & Kellman, 2015), mathematics (e.g. Kellman & Massey, 2013; Kellman, Massey, & Son, 2010; Landy & Goldstone, 2007), and other complex domains. Perceptual and adaptive learning modules (PALMs) accomplish this acceleration by using brief, interactive classification trials (e.g. Mettler & Kellman, 2014; Mettler, Massey, & Kellman, 2011). PALMs have only been created for visual stimuli thus far, which do not share the temporal nature of auditory stimuli. Can principles enhancing visual learning be extended to auditory stimuli, such as music?

B. Composer Style

Psychologists studying musical style perception have focused primarily on documenting human sensitivity to period styles (e.g. Hasenfus, Martindale & Birnbaum, 1983), leading to the development of hundreds of machine learning programs that simulate this impressive human ability (e.g. Widmer, 2005). However, while our human ability to recognize patterns in period styles has been well-documented, psychologists have done very little empirical work on whether humans have a similar sensitivity to composer styles, a topic that musicologists have long been interested in and discussed. Tyler (1946) played the three movements of a Mozart sonata and selected and played three movements each of trios by Beethoven and Schubert to students in a music appreciation course and found that students could distinguish between the composers better than chance, but she did not control for musical form (e.g. sonata or ballad) nor did she study learning. Crump (2002) chose J. S. Bach and Mozart because the composers were "stylistically distinct", and then only used the Goldberg variations for Bach and Minuets for Mozart. Participants and a machine learning algorithm successfully learned to distinguish these composers, but confounding musical form with composer limits the generalizability of these results. Due to the limitations of these studies, the empirical proof of composer style and of the learnability of composer style is uncertain.

C. Current Study

We developed a Composer PALM to apply perceptual learning technology to recognizing composers' styles, a challenging, abstract, perceptual learning task involving complex relations over time. We investigated whether (1) participants could learn composers' styles, and (2) whether the PALM framework could be successfully extended to a rich auditory domain.

II. METHOD

D. Participants

Forty-three (19 men, 24 women) undergraduate students at the University of California, Los Angeles participated for course credit in psychology or linguistics courses. Two participants were excluded for non-completion of the posttest. Included participants have learned to play an instrument for on average 7.17 years, and specifically, piano, for on average 6.88 years. Seventeen participants did not have any prior musical training.

E. Materials

1) *Music Stimuli.* There were clips from eight different composers from four different time periods (Renaissance: Byrd; Baroque: Bach, Handel, and Scarlatti; Romantic: Chopin, Mendelssohn, and Schumann; and post-Romantic: Debussy). The composers were chosen partially based on the size and availability of their musical work. The stimuli were 15-second clips of classical piano music collected from Youtube.com or the UCLA Music Library and were balanced with respect to stylistic features such as tempo and form. We chose to only use piano music so that instrumentation would not be confounded with composer or musical period. Thus, timbral information was relatively constant across recordings, periods, and composers. Participants, therefore, had to rely on more temporally extended patterns to learn to distinguish composers. We used only modern solo piano recordings (no harpsichord or four-hand recordings).

1) *Assessments.* The pretest and posttest assessments consisted of twenty-four clips from four composers trained in the Composer PALM (Bach, Handel, Schumann, and Chopin) and four untrained composers (Byrd, Scarlatti, Mendelssohn, Debussy).

Sixteen of the clips were from the trained composers (four each), and for each of these trained composers, about two clips were included in the PALM, and about two clips were new, novel clips that were not included in the PALM, to test both for learning in the PALM and for transfer. For each composer, the clips were divided among the composer's typical and atypical composition style. Experienced listeners determined typicality of composer style by listening extensively to all of the collected clips and subjectively categorizing them as typical or atypical of each composer's style.

The other eight clips were from the untrained composers (two each) - four from trained periods (Baroque and Romantic) and four from untrained periods (Renaissance and post-Romantic).

In each of 24 trials, the participants listened to a clip and chose who they think composed the music clip. They were given seven response options, clustered by period. Four of the options were the four PALM-trained composers, and the remaining three were "Other Baroque" (Scarlatti), "Other Romantic" (Mendelssohn), and "Other Period" (Byrd or Debussy) (see Figure 1).

There were three versions of the assessment (A, B, C), all containing unique clips. The amount of clips per composer in each version was held constant. The participants were randomly assigned versions in their pretest and posttest, without replacement, such that no participant received the same version in posttest as they did in pretest.

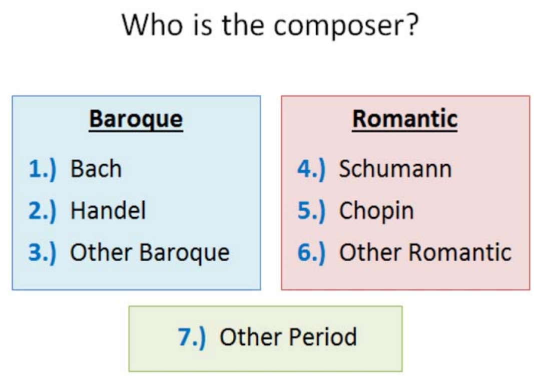


Figure 1. Response options for assessment. The "Other Baroque" composer was Scarlatti, the "Other Romantic" composer was Mendelssohn, and the "Other Period" composers were Byrd (Renaissance) and Debussy (post-Romantic).

2) *Intervention.* The Composer PALM consisted of 400 clips - 100 from each of four composers from two periods (Baroque period: Bach and Handel, Romantic period: Chopin, Schumann). Participants were presented with many trials, each containing one of the fifteen-second clips. In each trial, participants listened to a music clip and attempted to choose the correct composer from the four given composers (Bach, Handel, Chopin, and Schumann). If they answered correctly, the PALM showed that their answer was correct and let them proceed to the next trial (see Figure 1B). If their response was incorrect, the participants were shown the correct answer and listened to the remainder of the music clip (see Figure 1C).

For every twenty trials they completed, they were given a feedback summary on their average accuracy and average response time over the block of trials. Participants were also given feedback on mastery, using objective criteria. A participant reached mastery for a given composer when they correctly answered four of five consecutive trials of that composer, with a response time of less than 21s on those correct trials. Mastery level, the number of composers mastered, was indicated in the PALM by four circles on the bottom of the screen; for each composer mastered, one circle was filled in green. The module was delivered online, via a web browser.

F. Procedure

Participants took the pretest at the laboratory. Afterwards, they were instructed to undergo PALM training on their own time, using their own computer, for 45 minutes every day for seven days or until they completed the PALM by reaching mastery for all four of the composers. We sent daily emails reminding them to complete their training for the day as well as updating them on their progress. Following completion of the

PALM or a maximum of seven days of training, the participants returned to the laboratory to take the posttest.

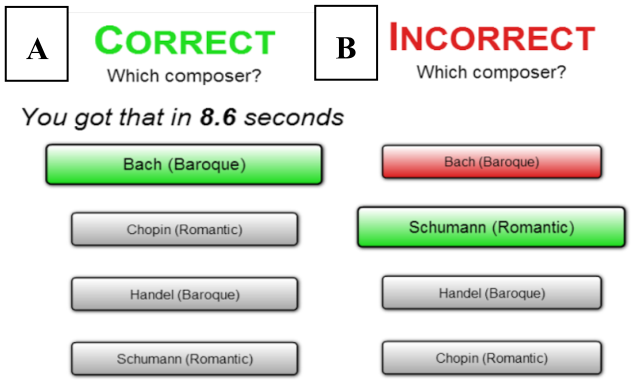


Figure 2. Screenshots of PALM trials showing feedback. Panel A shows a correctly answered PLM trial. Each trained composer is a response option, and their period is noted in parentheses. Feedback on accuracy was given on every trial. Panel B shows an incorrectly answered PLM trial. Participants saw which composer they had selected and which composed the music in the clip. After incorrect answers, participants listened to the rest of the clip.

G. Dependent Measures

This study used sensitivity as a dependent measure, with correction $1/(10n + 1)$, where n is the number of signal trials for hit rates and n is the number of noise trials for false alarm rates. For example, out of the twenty-four assessment trials, four were music clips by Bach. This means Bach had signal trials correction $1/(10(4) + 1) = 1/41$, and noise trials correction $1/(10(20) + 1)=1/201$.

III. RESULTS

H. PALM

On average, participants completed 482.10 (range: 94 - 979, $SE = 34.83$) trials in 4.25 (range: 0.58 - 13.32, $SE = 0.45$) hours over 6.95 (range: 1 - 16, $SE = 0.51$) PALM sessions before returning to the lab.

I. Pretest Assessment

At pretest, we evaluated participants' initial sensitivity to each composer and period style by testing whether their pattern of sensitivities differed from the hypothesis of no sensitivity ($d' = 0$). Specifically, we conducted a one-way repeated-measures analysis of variance (ANOVA) of Composer (Bach, Chopin, Handel, Mendelssohn, Scarlatti, Schumann, Other Period [Byrd+Debussy]) on sensitivity. Our custom hypothesis test of the insensitivity hypothesis confirmed that participants were insensitive to composer styles ($M = -0.11$, $SE = 0.17$) at pretest, $F(1,40) = 0.42$, $p = .52$, *partial-eta-squared* = 0.01. A second one-way repeated-measures ANOVA of Period (Baroque, Romantic, Other Period [Renaissance + post-Romantic]) on sensitivity with a custom hypothesis test confirmed that participants were also insensitive to period styles ($M = 0.21$, $SE = 0.15$), $F(1,40) = 2.08$, $p = .16$, *partial-eta-squared* = 0.05.

We also conducted a manipulation check of the three assessment versions via a mixed ANOVA of sensitivity with Assessment (A, B, C) as a between-subjects variable and Period as a within-subjects variable. There was no interaction, but there was a statistically significant main effect of Assessment such that performance on version A ($N = 12$, $M = -0.26$, $SE = 0.14$) was significantly worse than performance on versions B ($N = 12$, $M = 0.29$, $SE = 0.14$) and C ($N = 17$, $M = 0.24$, $SE = 0.12$), Pillai's Trace $F(1,40) = 4.91$, $p = .01$, *partial-eta-squared* = 0.21. Because of this performance difference, we analyzed improvement from pretest to posttest using analyses of covariance (ANCOVAs) of sensitivity change (d' change = posttest d' - pretest d'), controlling for pretest sensitivity.

J. Composers

One of our primary goals was to empirically investigate whether composer style could be identified and learned, by testing sensitivity to clips' composers. We conducted a custom hypothesis test to test the hypothesis of no sensitivity change (d' change = 0) in a repeated-measures ANCOVA for Composer (Bach, Chopin, Handel, Mendelssohn, Scarlatti, Schumann, Other Period [Byrd+Debussy]) on sensitivity change, covarying out pretest sensitivity. The custom hypothesis test confirmed that participants significantly improved their sensitivity to composers (see Figure 3) from pretest to posttest ($M = 0.61$, $SE = 0.21$), $F(1,33) = 8.29$, $p < .01$, *partial-eta-squared* = 0.20. We also found a significant main effect of Composer: participants improved on all composers (Bach $M = 0.26$, $SE = 0.14$; Chopin $M = 0.36$, $SE = 0.14$; Mendelssohn $M = 0.43$, $SE = 0.10$; Scarlatti $M = 0.29$, $SE = 0.10$; Schumann $M = 0.59$, $SE = 0.14$; Byrd and Debussy $M = 0.38$, SE = 0.12) except Handel ($M = -0.17$, $SE = 0.14$), Pillai's Trace $F(6,28) = 3.05$, $p = .02$, *partial-eta-squared* = 0.40.

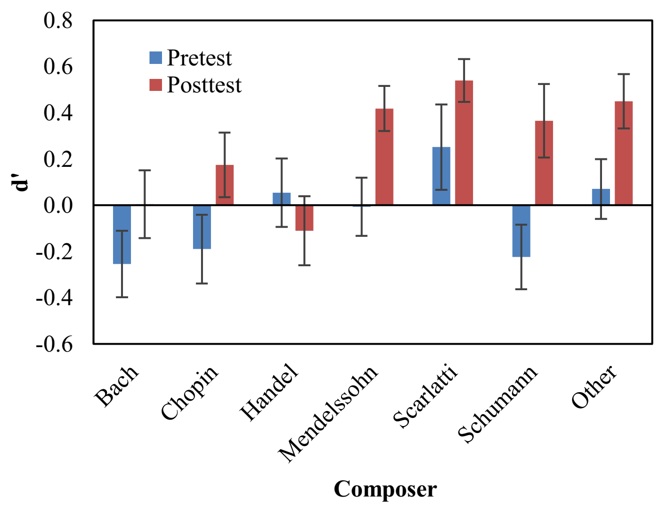


Figure 3. Participants' sensitivity at pretest and posttest for the seven possible composer options in the assessment. The trained composers (included in the PLM) were Bach, Chopin, Handel, and Schumann.

K. Periods

Given that our composers were clustered in the Baroque and Romantic periods and that sensitivity to period style has been well-documented (e.g. Hasenfus et al., 1983), we were interested to see if our participants would incidentally learn period style through PALM training on composers. We included a custom hypothesis test of the hypothesis of no change in sensitivity (d' change = 0) in a repeated-measures ANCOVA of Period (Baroque, Romantic, Other Period [Renaissance + post-Romantic]) on sensitivity change, covarying out pretest sensitivity to periods. The custom hypothesis test demonstrated that participants significantly improved their sensitivity to periods (M = 1.18, SE = 0.17) through incidental exposure during the PALM (see Figure 4), $F(1,37) = 40.06$, $p < .001$, *partial-eta-squared* = 0.52. The ANCOVA revealed no main effect of Period: participants performed equally well on the Baroque (M = 0.59, SE = 0.13) and Romantic (M = 0.78, SE = 0.15) periods and the combination of the Renaissance and post-Romantic periods (M = 0.38, SE = 0.12).

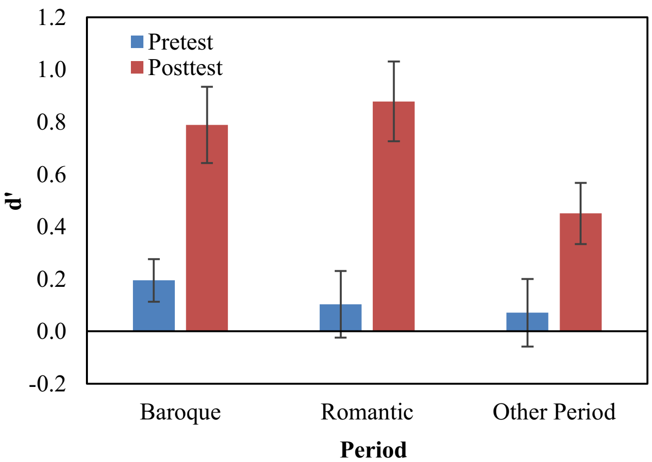


Figure 4. Participants' sensitivity at pretest and posttest for the two trained periods (Baroque and Romantic) and the untrained periods ("Other Period" = Renaissance and post-Romantic).

L. Familiarity

The goal of perceptual learning is not memorization of particular instances, but learning of structural regularities that facilitate accurate classification of new instances. If learning is primarily of this type, we would expect that posttest performance on familiar and novel examples should be similar. We included both familiar clips from the PALM and novel clips in each version of the assessment. To test for learning of the training music and for transfer to novel clips, and assess memorization of familiar clips as an explanation for our results, we compared sensitivity at posttest for familiar and novel clips. On period sensitivity at posttest, a two-way repeated-measures ANOVA of Period (Baroque, Romantic) and Familiarity (Familiar, Novel) confirmed no significant main effects or interactions. Participants showed no reliable differences for familiar Baroque (M = 1.14, SE = 0.26), novel Baroque (M = 0.80, SE = 0.20), familiar Romantic (M = 1.19, SE = 0.26), and novel Romantic (M = 0.96, SE = 0.19) clips. These results indicate that improvements were based on the pickup of relevant structure through perceptual learning rather than clip memorization.

Similarly, for composer sensitivity at posttest, we found little evidence of familiarity for each composer. We conducted a two-way repeated-measures ANOVA of Composer (Bach, Handel) and Familiarity (Familiar, Novel). We found no interaction or main effect of Composer, but we did find a main effect of Familiarity, such that sensitivity was higher on novel (M = 0.54, SE = 0.14) than familiar (M = -0.16, SE = 0.14) clips, $F(1,40) = 12.81$, $p < .001$, *partial-eta-squared* = 0.24. Participants were actually more sensitive to novel Bach (M = 0.38, SE = 0.18) and novel Handel (M = 0.72, SE = 0.22) clips than to familiar Bach (M = -0.23, SE = 0.18) and familiar Handel (M = -0.09, SE = 0.17) clips. This contradicts the explanation of participants memorizing clips because memorization predicts better performance on learned clips, not better performance on novel clips. There is no obvious explanation for the superiority of novel clips, but one possibility is that the clips used in training were somewhat less representative of the composer, or more difficult in some way, than the clips used in the assessments.

We could not include Chopin and Schumann in the above analysis because a programming error had caused assessment version B to have only novel clips (no clips included in the PALM) for Schumann, and assessment version C to have only novel clips for Chopin. To analyze familiarity for these composers, we conducted paired-samples t-tests on the data of participants not impacted by the error. For Chopin, participants (N = 27) were equally sensitive to familiar (M = 0.33, SE = 0.25) and novel (M = 0.38, SE = 0.30) clips, $t(26) = -0.12$, $p = .90$, *Cohen's d* = -0.02. For Schumann, participants (N = 26) were more sensitive to familiar (M = 0.91, SE = 0.32) than novel (M = 0.19, SE = 0.18) clips, $t(25) = 2.18$, $p = .04$, *Cohen's d* = 0.43. Of all the trained composers and periods, only Schumann lends any support to the explanation of familiarity or clip memorization. Overall, the familiarity results gave evidence that participants learned composers' and period styles.

IV. DISCUSSION

We used perceptual learning technology to investigate empirically human learning of composer styles. We found that participants successfully learned composer styles and incidentally learned period styles. Perceptual learning technology facilitated perceptual learning with music, complex auditory stimuli.

Gains in sensitivity were equal for trained and untrained composers. The improved sensitivity to untrained composers in trained periods (i.e. Baroque: Scarlatti, Romantic: Mendelssohn) and composers in untrained periods (Renaissance: Byrd, post-Romantic: Debussy) must then be due to learning of both trained composers' styles and of the Baroque and Romantic period styles.

Familiarity with the clips included in both the PALM and the assessments does not explain these results. Participants averaged 482.10 (SE = 34.83) trials in training with a 400-clip training set, so repeating clips, particularly clips included in the assessments, was a concern. Tests of sensitivity to familiar versus novel clips, however, generally revealed no effect of familiarity. For Bach and Handel, participants actually showed significantly greater sensitivity on novel clips!

In the absence of timbral cues and instrumentation, we found that participants could learn composer styles and period styles with PL technology. Because composers often favor particular forms, musical form is another potential confound, and prior work had confounded musical form and composer (Crump, 2002; Tyler, 1946). We selected a range of musical forms for each composer. Our participants demonstrated learning and sensitivity under more controlled and demanding conditions than those used in prior work. Thus, our results are strong empirical evidence for the existence and learnability of composer style in humans.

Perhaps psychologists have rarely previously studied composer style because of its complexity and difficulty to learn. Undergraduates who have learned to play an instrument (e.g. piano) for on average 7.17 years showed no sensitivity to composer or period styles at pretest; with PALM training they showed significant sensitivity to both composer styles and period styles. Seven years of musical training was insufficient to learn composer and period; but with only four and a half hours of PALM training, they gained expertise in these composer and period styles that would otherwise take years to learn. This is remarkable.

Our results demonstrate that perceptual learning technology is effective for learning auditory stimuli, even in complex and challenging domains. This suggests that perceptual learning interventions could accelerate learning in other complex auditory domains. Perhaps future work can use perceptual learning technology to investigate these domains, including important domains such as language.

ACKNOWLEDGMENT

The authors gratefully acknowledge support from NSF REESE award #110922 and the US Department of Education, Institute of Education Sciences (IES), Cognition and Student Learning (CASL) Program awards R305A120288 and R305H060070. We thank Timothy Burke, Joel Zucker, Rachel Older, and the Human Perception Lab research assistants for expert assistance.

REFERENCES

Crump, M. (2002, May). A principal components approach to the perception of musical style. In *Paper at the Banff Annual Seminar in Cognitive Science (BASICS),* Banff, Alberta.

Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York, NY: Appleton- Century-Crofts.

Hasenfus, N., Martindale, C., & Birnbaum, D. (1983). Psychological reality of cross-media artistic styles: Human perception and performance. *Journal of Experimental Psychology, 9*, 841–863.

Java, R. I., Kaminska, Z., & Gardiner, J. M. (1995). Recognition memory and awareness for famous and obscure musical themes. *European Journal of Cognitive Psychology, 7*, 41–53.

Karmarkar, U. R., & Buonomano, D. V. (2003). Temporal specificity of perceptual learning in an auditory discrimination task. *Learning & Memory*, *10*(2), 141-147.

Kellman, P. J., & Kaiser, M. K. (1994, October). Perceptual learning modules in flight training. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 38, No. 18, pp. 1183-1187). SAGE Publications.

Kellman, P. J., Massey, C. M., & Son, J. Y. (2010). Perceptual learning modules in mathematics: Enhancing students' pattern recognition, structure extraction, and fluency. *Topics in Cognitive Science, 2*(2), 285–305.

Mettler, E., & Kellman, P. J. (2014). Adaptive response-time-based category sequencing in perceptual learning. *Vision Research*, *99*, 111-123.

Mettler, E., Massey, C., & Kellman, P. J. (2011) Improving adaptive learning technology through the use of response times. In L. Carlson, C. Holscher, & T. Shipley (Eds.), Proceedings of the 33rd Annual Conference of the Cognitive Science Society (pp. 2532-2537). Boston, MA: Cognitive Science Society.

Sabin, A. T., Eddins, D. A., & Wright, B. A. (2012). Perceptual learning evidence for tuning to spectrotemporal modulation in the human auditory system. *The Journal of Neuroscience*, *32*(19), 6542-6549.

Sagi, D. (2011). Perceptual learning in vision research. *Vision Research*, *51*(13), 1552-1566.

Song, Y., Peng, D., Lu, C., Liu, C., Li, X., Liu, P., et al. (2007). An event-related potential study on perceptual learning in grating orientation discrimination. *Neuroreport*, *18*(9), 945–948.

Thai, K. P., Krasne, S., & Kellman, P. J. (2015). Adaptive perceptual learning in electrocardiography: The synergy of passive and active classification. *Proceedings of the 37th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Tyler, L. E. (1946). An exploratory study of discrimination of composer style. *The Journal of General Psychology*, *34*(2), 153-163.

Widmer, G. (2005). Studying a creative act with computers: Music performance studies with automated discovery methods. *Musicae Scientiae*, *9*(1), 11-30.

Detecting offset asynchrony for pairs of tones with musically-relevant durations

Joshua Shive[1,3], David Heise[2,3], Emily Ector-Volman[1], Justin Manson[1], Shereece Steinmann[1]
[1]Department of Psychology, Tennessee State University, USA
[2]Department of Computer Science, Technology & Mathematics, Lincoln University, USA
[3]CRoMA (Computational Research on Music & Audio) Team of Interdisciplinary Collaborators (CRoMA-TIC), Lincoln University, USA

Offset detection has received relatively little attention in the field of auditory perception compared to onset detection. Previous studies of offset detection thresholds for musical elements have examined *harmonic* offsets within a complex tone (Zera & Green, 1993), as well as *temporal* offsets of pairs of tones (Pastore, 1983). In a temporal offset detection task, two tones with identical onsets but slightly different offsets are played through headphones, and a listener is asked to determine which tone ended first. By varying the offset asynchrony of the tones across trials, one can estimate the listener's temporal offset judgment (TOJ) threshold: the minimum offset asynchrony required to detect that the tones end at different times. The durations of the tones used in these studies are typically very short (e.g., ranging in duration from 10 to 300 ms in Pastore, 1983). In contrast, notes in a musical piece, whether played by an individual instrument or an instrumental section, can last for several seconds. Heise, Hardy, and Long (2015) reported a TOJ threshold of approximately 100 ms for tones lasting 1500-2000 ms, but this result warrants validation as it is much higher than reported by Pastore and assumes 100% detection (vs. 70.7% by Pastore). Our current project estimates TOJ thresholds for a set of tones with durations similar to what listeners might encounter in a piece of music played at a moderate tempo (120 bpm). We are conducting a psychophysical experiment to estimate TOJ thresholds for two inharmonic tones (256 Hz and 400 Hz) played at 125, 250, 500, 1000, 1500, and 2000 ms (corresponding to sixteenth, eighth, quarter, half, dotted-half, and whole notes, respectively). We predict that TOJ thresholds will increase logarithmically as note duration increases. A function that relates note duration and TOJ threshold will be useful for designing a musical transcription system or other sound event detection algorithm that matches human perception.

Examining the cognitive and perceptual differences between musicians and non-musicians: A meta-analysis

Katherine Thompson[1], Mark A. Schmuckler[1], Konstantine Zakzanis[1]
[1]Department of Psychology, University of Toronto Scarborough, Toronto, Ontario, Canada

Music training has been associated with improvements in both musical and non-musical cognitive domains, as reported by a number of studies. Conflicting findings, however, make it difficult to conclude exactly in what fashion musicians outperform non-musicians through systematic reviews alone. To address this question, two meta-analyses were conducted to delineate whether musicians and non-musicians differ across a range of cognitive and music processing domains, and to determine the magnitude of such differences. Moreover, potential covariates (derived from existing literature) were included as moderator variables. The most common of such covariates was the amount of training that musicians have received; thus, it was hypothesized that years of training would predict performance across domains. Additional variables believed to contribute to differences between musicians included age of training onset, type of training, and so on, and were similarly employed as moderator variables. As predicted, musicians significantly differed from non-musicians on all music processing domains. Musicians also differed on many cognitive domains: Construction and Motor Performance (d=.68), Executive Functions (d=.65), Memory (d=.29), Neuropsychological Assessment (d=.31), Orientation and Attention (d=.40), Perception (d=.35), and Verbal Functions and Academic Skills (d=.43); but not Concept Formation and Reasoning (d=.183). Of more surprise was the finding from meta-regression analyses that years of training was not related to performance in most of the cognitive and music processing domains. In fact, years of training predicted superior performance only in Construction and Motor Performance and Temporal Perception (p<.05). This pattern indicates a surprising finding that, although musicians and non-musicians clearly differ across a range of musical and cognitive processing dimensions, the amount of training does not explain differences. These results highlight the need for a standardized protocol for defining musician samples.

An exploration of psychoacoustic factors influencing the perception of triangular sound shapes

Sven-Amin Lembke[1]
[1]Music, Technology and Innovation Research Centre, De Montfort University, Leicester, UK

In electroacoustic music, the way in which sound spectra evolve over time often resembles geometric shapes (e.g., Smalley's 'spectromorphology'). For instance, such shapes may delineate a spectrotemporal evolution in which two sides of a triangle diverge in frequency over time. Although the perceptual clarity of the shape could be implied by the diverging sides' trajectories, the spectral content enclosed therein could also bear some morphological significance, suggesting that sound shapes could involve several perceptual aspects. This study explored the acoustic factors influencing the perception of sound shapes and how they contribute to several perceptual attributes. Triangular sound shapes, constant in duration and frequency extent, were studied as a function of principles known from auditory scene analysis and different scalings of frequency and amplitude. A listening experiment employed stimuli created by granular synthesis, which allowed the control over the density of sinusoidal grains. Two independent variables varied the proximity among sound grains along frequency and time, respectively. Two remaining independent variables involved different frequency and amplitude scalings, based on either linear or psychoacoustic functions. The dependent variables considered several ratings, assessing the perceived degree of clarity, opacity or balance of the implied shape and the sound's degree of heterogeneity. Results from a pilot study suggest that all investigated factors influenced perceptual aspects of sound shape. Clarity of shape appeared to be largely unaffected by granular density, whereas density appeared to influence other perceptual aspects. These preliminary trends require further confirmation in the forthcoming months, which may allow associating acoustic factors with the perceptual aspects by means of multivariate regression. Further insight into the underlying factors will inform more complex perceptual studies on electroacoustic music.

Knowledge-Based Expectation Effects on Pitch Perception: A cross-cultural Behavioral and ERP Investigation

Elif Canseza Kaplan[1], Z. Funda Yazıcı[2], Esra Mungan[3]
[1]Cognitive Science Program, Boğaziçi University, Istanbul, Turkey
[2]Musicology and Music Theory Department, Istanbul Technical University, Istanbul, Turkey
[3]Psychology Department, Boğaziçi University, Istanbul, Turkey

Earlier studies have shown that harmonic (e. g., Bigand & Pineau, 1997) as well as tonal expectations (e. g., Marmel et al., 2008; Marmel et al., 2011) influence pitch processing. Moreover, we know that different musical systems have different pitch representations. Turkish makam music, for instance, is a microtonal musical system that divides an octave into 24 or 36 instead of seven intervals. Hence one should expect listeners from that culture to have a different categorical representation of pitch compared to Western listeners, which in turn might influence their sensitivity to minute pitch deviations. The current study aims to test the influence of both tonal expectation and musical culture on pitch perception. In our experiment, Turkish makam musicians, Turkish nonmusicians, and Western listeners are asked to judge whether a final note within a Turkish makam or Western melodic sequence is in- or out-of-tune. In some cases, the final note completes the melodic sequence ("full cadence"), in some it does not ("half cadence"). We expect faster and more accurate out-of-tune judgments for notes that complete a cadence (cf. Bigand & Pineau, 1997). Furthermore, due to its microtonal system we expect listeners encultured and entrained within Turkish makam music to be more accurate in judging pitch deviation due to potentially possessing frequency-wise more narrow pitch category representations. It is likely that Western listeners will show a response bias effect in the makam music context compared to the Western music context. Results will be discussed within the framework of top-down and attention-related processes. The study also plans to explore ERP correlates of expectation effects on pitch accuracy judgments.

Influence of Information: How Different Modes of Writing about Music Shape Music Appreciation Processes

Timo Fischinger,[*1] Michaela Kaufmann,[*2] Wolff Schlotz[*3]

[*] *Music Department, Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, Germany*

[1]timo.fischinger@aesthetics.mpg.de, [2]michaela.kaufmann@aesthetics.mpg.de, [3]wolff.schlotz@aesthetics.mpg.de

I. BACKGROUND

In every concert or while listening to a CD we are familiarized to the music heard by informative texts, such as program notes and CD booklets. Already with its integration into nineteenth century concert life, informative texts had three central functions: 1) providing knowledge of the work in the more narrow sense, e.g. example information on the history of a work's creation or a composer's biography; 2) providing knowledge regarding classification and valuation, e.g. the significance of a work's compositional and cultural history; 3) providing knowledge of interpretation and reception.

II. AIMS

Assuming that these types knowledge provided by texts about music may have an effect, first, on the listeners' aesthetic judgments, and second, on the listeners' perception of the music, we designed two listening experiments adapting the framing paradigm to test these assumptions.

III. METHOD

For both experiments participants interested in classical music were recruited (same group of participants for both experiments): N = 170; 56% female; mean age: 45.84 (17.30) years (min = 19, max = 80); general musical sophistication (Gold-MSI): 82.68 (18.22); all from the greater area of Frankfurt. In the first experiment, participants were asked to listen three times to identical excerpts of the third movement of the 1st symphony in C minor op. 68 by Brahms in a within-subject design. Prior to listening, diverging descriptions of existing recordings either with the conductors 1) Mariss Jansons, 2) Christian Thielemann, or 3) Daniel Barenboim were presented for each excerpt. In the second experiment participants were asked to listen to the first movement of the symphony in E flat major for 2 horns, 2 oboes and strings by J. Mysliveček (1737–1781) played by Concerto Köln with Werner Ehrhardt. Half of the group was told that they would be listening to W. A. Mozart whereas the other half was told they would be listening to Mysliveček. Apart from that, the composer's names were combined with texts containing either an emotional-expressive or structural description resulting in a 2x2 between groups design. In both experiments, participants rated the music in the domains of the perception of musical structure, sound qualities, musical and emotional expressiveness, as well as liking.

IV. RESULTS

Preliminary analyses suggest that the manipulation of information provided influenced the way the music was perceived and evaluated. Ratings of the depended variable liking were significantly higher in the second experiment if the music was preceded by texts with an emotional-expressive compared to a structural description of the music [main effect of text: $F(1, 166) = 11.40$, $p < .001$, $\eta^2 = .064$], while no effect of composer and/or interaction between the two variables text and composer as well as of musical expertise (Gold-MSI) was found. Further analyses of the semantic differential data as well as the data from the first experiment will be reported at the conference.

V. CONCLUSION/DISCUSSION

Contextualized listening not only influences the listener's subjective attitude, as studies on the prestige effect have shown extensively, but in addition, different modes of writing about music seem to affect liking of the music heard. Likewise, the perception of aspects pointing at the music's meaning, such as emotional expressiveness, are more likely to be influenced by different modes of writing about music than the perception of aspects on a more technical level, such as musical structure.

Cognitive and affective reactions to repeating a piece on a concert program

Andrea R. Halpern[1], John Sloboda[2], Chloe H.K. Chan[3], and Daniel Müllensiefen[3]
[1]Psychology Department, Bucknell University, Lewisburg, PA, USA
[2]Guildhall School of Music and Drama, London, UK
[3]Psychology Department, Goldsmiths, University of London, London, UK

Repetition of a piece on a concert program is a well-established, but uncommon performance practice. Musicians have presumed that repetition benefits audience enjoyment and understanding but no research has examined this. In two naturalistic and one lab study, we examined audience reaction to live performances of two contemporary pieces played by the same ensemble on the same program. In line with mere exposure research, we expected liking to increase upon second hearing, although using contemporary music and nonidentical performances might work against an increase. We expected increases in perceived understanding, predicted, and actual memory, given a second learning trial. Study 1 comprised two nonconsecutive performances of Varèse's *Ionisation*; the audience was pre-informed of the repetition. Study 2 comprised two pairs of original student compositions: one pair was played back-to-back and the other was not; the audience was informed of the repetition just before the second playing. In Study 3, the fourth (of four) pieces was a repeat of the second, recorded from live repeated performance; half the listeners were pre-informed of the repetition. In all studies, we asked listeners to rate their enjoyment and willingness to hear the piece again (Affective), and perceived understanding and predicted memory of the piece (Cognitive). In Study 3, we assessed immediate recognition memory of each excerpt. In all studies, Cognitive variables increased significantly. Affective reaction also increased except for *Ionisation* in live listening; however that piece was well liked at first hearing. Memory performance was low and not related to predicted memory, nor increased after a second hearing, although memory scores were correlated in the two hearings of the same (but not different) piece. Being informed or not had no systematic effect. Audience and performer reaction was mixed. We discuss the challenges and opportunities in studying this interesting performance practice.

Evaluating Recorded Performance: An analysis of critics' judgements of Beethoven piano sonata recordings

Elena Alessandri, Victoria J. Williamson, Hubert Eiholzer, Aaron Williamon

***Abstract*—**Reviews of Beethoven's piano sonata recordings (N = 845) published in the *Gramophone* (1923-2010) were collated and analyzed using a combination of data reduction and thematic techniques in order to identify performance features and extra-performance elements that critics discuss and reasons adduced to support their value judgements. This led to a novel descriptive model of recorded performance critical review. The model captures four activities in critical writing – evaluation, descriptive judgement, factual information and meta-criticism – and seven basic evaluation criteria on the aesthetic and achievement value of performance, plus two recording-specific criteria: live-performance impact and collectability. Critical review emerges as a dense form of writing open to diverse analytical approaches. Findings inform current discourses in aesthetics and open new perspectives for empirical music research. In particular, they stress the importance of the notion of personal achievement in written performance evaluation and the role of critics as mediators in the recording market.

***Keywords*—**Aesthetic response, Beethoven, music criticism, recorded performance.

I. Introduction

What makes for a great performance? This question has attracted the interest of musicians, philosophers, and scientists for centuries, back to Hume's debate on the nature of aesthetic judgements [1] and Helmholtz's studies on physiological responses to music [2]. In recent decades studies have identified both formal properties (e.g. balance between stimulus simplicity and complexity) and circumstantial factors (e.g., listener's personality, enculturation, and familiarity) that drive aesthetic response [3]-[5] as well as the underlying cognitive mechanisms [6].

A corpus of research has focused on performance grading in educational contexts: studies examined consistency and reliability [7]-[10] and identified musical and extra-musical factors that enter listeners' evaluation, including performer's gestures and physical appearance [11]-[17].

E. Alessandri is with the School of Music, Lucerne University of Applied Sciences and Arts, 6003, Switzerland (phone: +41-412-492642; e-mail: elena.alessandri@hslu.ch).

V. J. Williamson is with the Department of Music, University of Sheffield, UK (e-mail: v.williamson@sheffield.ac.uk).

H. Eiholzer is with the Department of Research and Development, Conservatory of Southern Switzerland, Lugano, Switzerland (e-mail: hubert.eiholzer@conservatorio.ch).

A. Williamon is with the Centre for Performance Science, Royal College of Music, London, UK (e-mail: aaron.williamon@rcm.ac.uk).

These efforts have improved understanding of the phenomena underpinning the listening experience and the determinants of performance appreciation. However, there remains no consensus on the nature of criteria that may reliably drive spontaneous evaluation in music listening. In order to offer a new perspective on this debate, [18], [7] and more recently [19] and [9] have all made calls for structured, inductive qualitative research to explore evaluation by experts in diverse ecologically valid contexts.

In this study we addressed this call directly through a systematic examination of expert critics' reviews of recorded performance. Critical review of recorded performance is a popular professional and commercial form of written response to music. Recording reviews are published in paper and digital formats and are relevant to listeners' choices and musicians' careers. However, despite the availability of material and its relevance to the musical praxis, there has been little structured enquiry on how music critics make sense of their listening experiences [20], [21]. In particular, no research has explored the criteria critics use in their evaluations of recorded performance.

Our aim was to offer the first structured investigation of a vast corpus of recording critical reviews. By looking at what critics write, what aspects of the performance and the end-product recording they seek out for critical attention, and how they structure their judgements, we aimed to understand how expert listeners conceptualize and evaluate performance. Our two research questions were: what do critics write about when they review recorded performance? What reasons do they adduce to support their value judgements?

II. Method

We developed a hybrid qualitative/quantitative analysis design, based on the Applied Thematic Analysis framework [22]. Our protocol (Figure 1) featured successive steps of quantification and data reduction techniques followed by in-depth thematic analyses.

The protocol was semi-structured: results of previous analyses informed and shaped subsequent steps. This allowed a thorough examination of the review material from the metadata, word-stem, clause, and multiple clause analysis levels.

A. Preparation of material

We collected reviews of Beethoven's piano sonata recordings (N = 845) published in the monthly British magazine *Gramophone* between April 1923 and September 2010 (1'050 issues). Critical text (i.e. excluding titles, technical information, and critic's name) was extracted and metadata information on issue, critic, pianist and sonata reviewed, label, presence of comparison between pianists, and release status (new recording or re-issue) were collated into a searchable database.

B. Overview and data reduction

Next, descriptive and inferential statistics were run on the metadata to obtain an overview of critics, repertoire, and pianists involved [23]. A series of quantitative/qualitative data reduction techniques were applied to gain insights into the textual content of reviews. These included: (i) a thick-grained thematic analysis of a sample from the corpus (n = 67 reviews); (ii) a comparison of word-stem patterns between critics and between periods of publication (ReadMe estimations [24]); (iii) a qualitative analysis of critics' vocabulary with terms grouped into semantic categories followed by (iv) a comparison of the relative use of semantic categories between critics and periods of publication (LIWC analysis [25]), and (v) a word-in-context [26] analysis to disambiguate critics' use of the term 'expression' and its correlates. Details of these analyses are reported in [27] and [28]. From the findings of these preliminary analyses a representative selection of reviews (N = 100) was produced, to be used in the subsequent thematic analyses.

C. Core thematic analyses

Three thematic analyses were run on the 100 selected reviews. Analysis I mapped the different performance features discussed in critical review [27]. Analysis II identified other recording elements discussed (e.g. price and recording quality) and examined how they entered into critics' final judgements. Analysis III focused on the overall valence expressed by critics statements and identified evaluation criteria underpinning critics' judgements [29].

A double-coder analysis protocol was developed for Analyses I and II based on [30] and [31]. This encompassed a double-coder line-by-line open coding of a subset of the texts, with codes discussed between researchers and structured into a codebook. This was followed by the main researcher's application of the codebook to the whole dataset (segmentation at clause level) and by an iterative process of analysis and comparison of quotes for each code to clarify subcategories and relationships between themes.

For Analysis III the two researchers coded the whole dataset separately and then came together to agree on whether each statement from the corpus presented with a positive, negative, or neutral valence. Valence loaded statements were then cross-analyzed with the descriptors found in Analyses I and II to identify the main and sub-criteria that were used to support critics' evaluative judgements.

D. New model development

Upon completion of the three analyses, all the findings were compared and combined into new descriptive models that address directly our research questions; what are the key components of critical review content and how they are evaluated by critics.

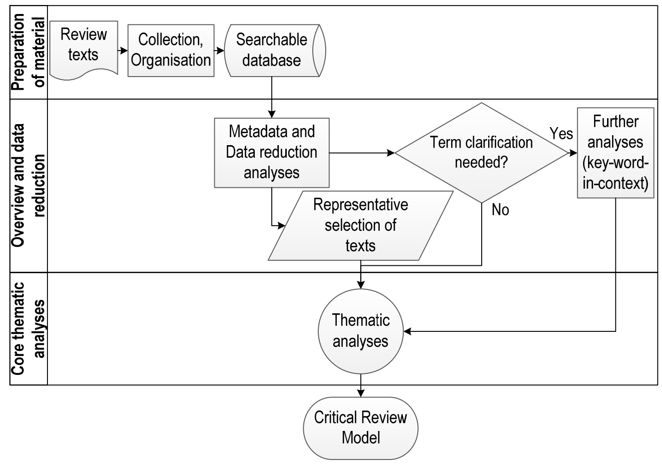


Fig. 1 Schematic flowchart representation of the analysis protocol applied in the present research.

III. Results

The initial 845 *Gramophone* reviews encompassed discussion of 640 performances and 205 re-issued recordings performed by 216 pianists and reviewed by 52 critics. Comparisons between recordings were found in 53% of texts and reviews polarized around a few performers: 17 out of 216 pianists covered 51.95% of the review corpus.

Ten critics with a mean activity period of 21.32 years wrote 62.72% of reviews. Over the 87 years, the use of semantic categories (words per review) varied more strongly between critics than between periods of publication (averaged Kruskal-Wallis across semantic categories: H_{10} = 52.30, p = .037 between critics, H_6 = 25.56, p = .086 between decades).

As part of the thematic analyses, 10 reviews were selected for each of the 10 most prolific critics in the original sample. These 100 reviews (35,753 words) were published between 1934 and 2010, and entailed discussion of at least six recordings for each of Beethoven's 32 piano sonatas, as performed by a total of 56 different pianists.

From the combined findings of Analyses I, II, and III, four supervenient theme families emerged, with 14 main themes and 55 sub-themes. The first three supervenient theme families, **Descriptive Judgement**, **Factual Information**, and **Meta-Criticism**, map out the different topics discussed in reviews (research question one: Figure 2). The fourth superordinate theme family, **Evaluative Judgement**, summarizes the basic evaluation criteria underpinning critics' statements (research question two: Figure 3).

A. The content of recording critical review

Figure 2 visualizes the thematic outcomes of Analyses I and II. For the sake of brevity the model summarizes the emergent themes down to the fourth hierarchical level. Levels five and six are reported in detail in [27].

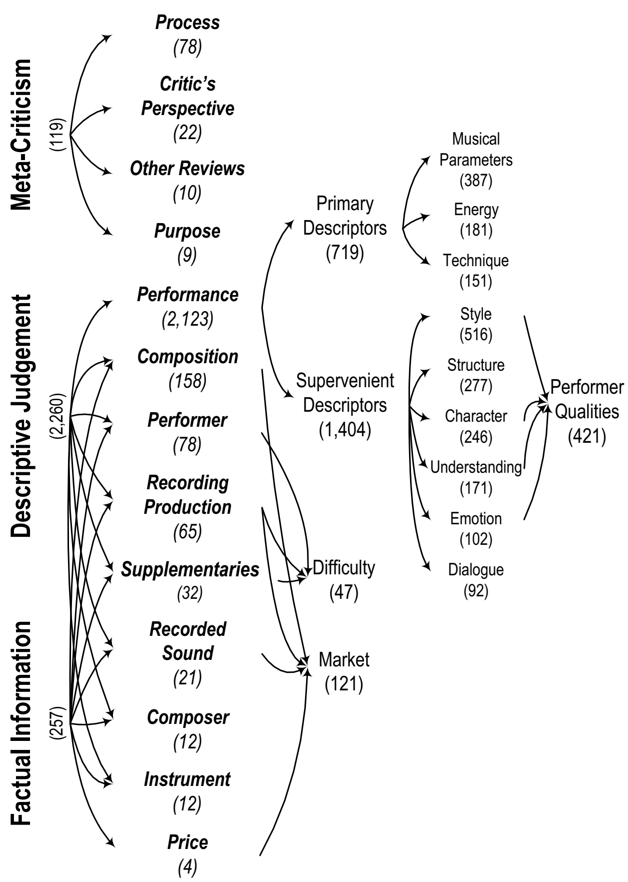


Fig. 2 Summary visualisation of the themes that emerged from Analyses I and II. Superordinate theme families are located in the left-hand side. Themes are visualised hierarchically moving from left to right down to the fourth hierarchical level. In parentheses under each theme name is the number of times the theme was coded in the texts. Each time a sub-theme was coded, the relevant higher-order themes were coded as well.

The largest theme family in reviews entailed characterizations of the recorded performance that require the critic's opinion or subjective conclusion (**Descriptive Judgement**, 2,260). The largest and most varied group of **Descriptive Judgements** were characterizations of the ***Performance*** (2,123). This group entailed comments on the musical sound, level of energy, and mechanics of delivery (Primary Descriptors, 719) as well as higher-order impressions of the performance in terms of artistic style, structure, dialogue, character, emotion, and understanding (Supervenient Descriptors, 1,404). Among Supervenient Descriptors, a group of statements characterized the performance focusing on presumed Performer Qualities (a performer's understanding, affective states, moral qualities, or intentionality: 421). The construct of musical expression emerged in reviews as multi-layered: the term 'express' and its correlates were used in relation to Primary and Supervenient Descriptors to indicate the way performers use their performance options (e.g., timing or dynamics), the portrayal of music structural patterns, or the outer manifestation of (defined or undefined) inner states.

Besides ***Performance***, reviews entailed, on average, discussion of at least three other recording elements. These were mainly characterizations of the ***Composition*** (80) in terms of character, structure, or compositional style plus a few considerations on the content of sleeve notes (***Supplementaries***, 12), character of the ***Recorded Sound*** (10) or ***Performer's*** general artistic or personal traits (10). A subgroup within these comments focused on challenges faced by the people involved in the recording production (composer, performer, engineers: Difficulty, 47). The relative frequency with which the different **Descriptive Judgements** were used in reviews was highly consistent between the ten critics (Cronbach's α = .986 for performance statements and α = .962 for extra-performance statements).

Besides **Descriptive Judgements**, reviews encompassed comments on facts (historical or current) that describe or contextualize the recorded performance (**Factual Information**, 257). Most **Factual Information** was about the ***Composition*** performed (78), the ***Performer*** (68), and the ***Recording Production*** (65). A subgroup of **Factual Information** discussed the recording in the context of the wider music market (Market, 121), for instance indicating the availability of other recordings of the same piece(s) or by the same pianist.

Finally, **Meta-Criticism** encompassed reflections on the ***Process*** of review writing (78), on the reviewer's past experiences, preferences, or dispositions (***Critic's Perspective***, 22), on statements by colleagues (***Other Reviews***, 10), and on the function of critical review as consumers' guidance (***Purpose***, 9). These comments emerged as relevant to the interpretation of **Evaluative Judgements**, as they aimed to alert the reader to possible critic biases or hinted at how readers should use the review when deciding what to buy.

B. Evaluation in Recording Critical Review

Figure 3 illustrates the outcomes of Analysis III. Extra analyses were performed on the ***Performance*** theme due to its density (6.66 codes/clause). These are reported in [29]. For the purpose of this article, the model serves to illustrate how this much larger theme can be conceptualized alongside the other elements.

Evaluative Judgements entailed comments about the worth, merit, importance or usefulness of the recorded performance as a whole (***End-Product***, 28) or of its elements. This was the most pervasive form of critical activity – the only one found in each of the 100 reviews – and the largest single theme arising from both the analysis of ***Performance*** (1,502) and extra-performance (346) related statements.

In particular, Analysis III showed that ***Performance*** evaluations in reviews were expressed both explicitly through

Evaluative Judgements and implicitly through **Descriptors** that also entailed an evaluative connotation (e.g., ‘nimble fingers’ vs. ‘overtaxed fingers’). In total, 87.57% of ***Performance*** statements in reviews were valence loaded. Comments on Difficulty (Figure 2) also shaped evaluations, emphasizing the value of a performative outcome as the performer’s, composer’s, or engineer’s achievement.

The systematic examination of the relationship between descriptors and valence loaded statements showed that critical evaluations can be explained through a small number of basic criteria used reliably by all critics (Kendall's coefficient of concordance $W = 0.81$, $p < 0.001$):

The first seven criteria were linked to the *Suitability* of all the criteria given the musical context, to the Aesthetic value of the recorded performance (*Intensity*, *Coherence*, and *Complexity*) and to its Achievement value (*Sureness*, *Comprehension*, and *Endeavour*). A tension emerged between these seven criteria: critics discussed them as if they were interdependent, so that an increase in one criterion could add or mar other criteria (Tightrope).

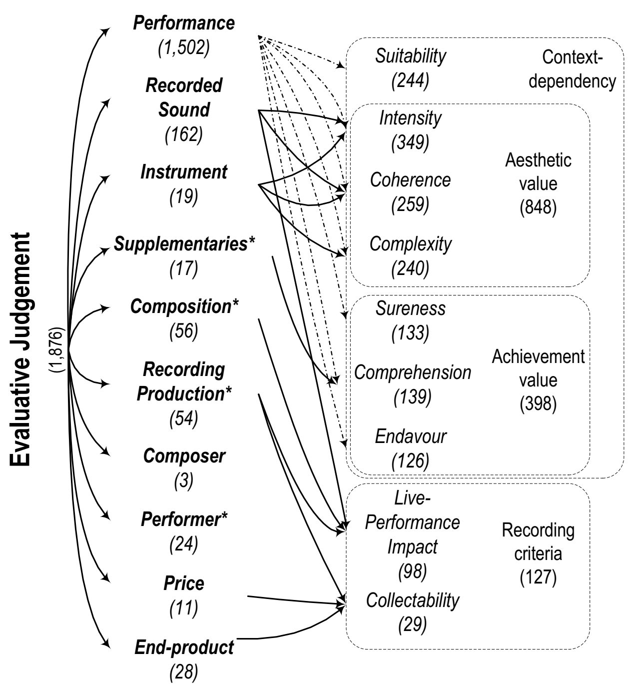


Fig. 3 Thematic outcome of Analysis III. Themes are visualised hierarchically moving from left to right with evaluation criteria emergent across the different themes clustered on the right side of the model. The solid arrows represent related analyses while the dotted arrows emanating from ***Performance*** visualize a process where analyses had to be broken down due to the size and density of the ***Performance*** theme. Asterisks represent themes were comments could be addressed to different target audiences (labels, performers).

In addition to the seven criteria named above, two recording-specific criteria emerged, linked to the live performance-like experience the recording offers (*Live-Performance Impact*) and to its value as an historical document or collector’s item, linked to the semi-permanent nature of the recording object (*Collectability*).

14.65% of performance and 6.42% of extra-performance **Evaluative Judgements** were comparisons, in which the recorded performance was discussed against one or more other recordings. Reviews were also characterized by a mixture of absolute and relative (taste-dependent) **Evaluative Judgements**, with taste-dependent judgements present in 47% of reviews. Finally, a sub-group of **Evaluative Judgements** (24) addressed the record producers (labels, performers), criticizing their course of action or expressing a desire for a certain product. These comments fell within the themes ***Supplementaries***, ***Composition***, ***Recording Production***, and ***Performer*** and are heightened with an asterisk in Figure 3.

IV. Discussion

We conducted the largest analysis to date on a sizeable corpus of piano recording reviews in order to outline the performance features and other recording elements that critics discuss and the criteria they adduce to support their value judgements. The final result is two descriptive models that capture critics’ listening focus and what they think makes for a great recorded performance.

The overall emergent picture describes the content of recorded performance critical review in terms of **Descriptive** and **Evaluative Judgements** of the ***Performance*** plus eight other recording elements, particularly the ***Composition***, the ***Recorded Sound***, and the ***Recording Production***. Critics also provided **Factual Information** on the recording itself and on the surrounding recording market, as well as meta-reflections on the process and nature of critical writing (**Meta-Criticism**).

Evaluative Judgements emerged as the primary critical activity in reviews, one that is embedded in a rich variety of ***Performance*** descriptors alongside characterizations of the ***Composition*** performed. This finding supports the view of music criticism as reasoned evaluation [32]. ***Performance*** **Descriptors** used by critics overlapped partially with those commonly used in research [10], [8] and educational contexts [11]. Here however, musical ‘expression’ did not emerge as a cohesive construct, but rather as a fluid and multi-layered notion linked to different Primary and Supervenient Descriptors. The ease with which critics moved from a physical (use of musical parameters) to a psychological (communication of inner states) notion of ‘expression’, calls for caution in the use of this term in evaluation studies.

Despite the variety of **Descriptive** and **Evaluative Judgements**, appraisals could be explained in terms of a few criteria: *Intensity*, *Coherence*, and *Complexity* were linked to the Aesthetic value of the recorded performance in line with [33], [34] and [35] theories of aesthetic appreciation. Three more criteria – *Sureness, Comprehension*, and *Endeavour* – focused on Achievement value. Together with the Performer Qualities descriptions and comments on Difficulty, these findings support the intentionalist view [36], [37] and the ‘success value’ theory [32] and emphasize the importance of perceiving the final product as a joint outcome between the performer’s, composer’s and engineers’ achievements. Future research should explore the extent to which this notion of achievement could be efficiently integrated in educational assessment schemes.

The different descriptors and criteria we have identified were used reliably by all our critics across almost 90 years of published review. However the criterion *Suitability*, the tension between aesthetic and achievement criteria (Tightrope), and the presence of comparative and taste-dependent judgements, all emphasize the context-dependency of performance evaluations in line with Sibley's [38] context-aware generalism. This resonates with interactionist theories on the aesthetic experience such as the 'uniformity-in-variety' theory [39] and the processing fluency framework [6], and suggests that future developments of assessment protocols should account for the complex and dynamic interaction between criteria.

In addition to the Aesthetic and Achievement values, two recording-specific criteria emerged that focus on the recording as a collectable and on its ability to offer a listening experience akin to a concert (*Collectability* and *Live-Performance Impact*). These two criteria and the weight given in reviews to extra-performance elements, identify the recorded performance as a unique artistic product, distinct from live performance [40], [41]. In particular, they point at what [42] defined as the tension between 'creating virtual worlds' and 'capturing performances' and support the view that recorded performance judgements need to account for the way in which the sound "got onto the disc" [40].

Finally, the polarization of reviews around a few pianists and performances, the use of comparative judgements, comments on ***Price*** and Market, and recommendations addressed to external agents, all emphasize the critics' role as mediators between producers and consumers. They aim to act as a guide for consumers when deciding what product to buy [43]-[45]. The next step for this research will explore the extent to which this purpose is consciously intentional and its impact on consumer choices.

V. Conclusions

We analyzed a vast corpus of critical review to map out what expert critics write about and how they justify their evaluations of recorded performance. Despite the focused nature of the data source (*Gramophone*), repertoire (Beethoven's 32 piano sonatas), and format (recorded performance) the material has provided multiple insights relevant to research and musical practice.

The findings support but also challenge theories on aesthetic appreciation; they provide musicians with a detailed account of critics' focus in their assessment of Beethoven's sonatas – essential in any pianists' repertoire – and offer for the first time two models of critical evaluation that account for the specific nature of recorded performance as an artistic product.

A main aim of this research was to offer a new perspective on the performance evaluation discourse by looking at what we can learn from expert critics' writings. Performance critical review emerged as a rich source material apt to systematic investigation of both content and purpose: the present methods and models offer solid empirical grounds for future comparative investigations of other forms, genre, corpora or aspects of music critical review.

Acknowledgment

We thank the *Gramophone* for permission to study their archive for this research and the Institute for Quantitative Social Science at Harvard University for permission to use their text analysis algorithm. We also thank Alessandro Cervino for his assistance in collecting data and Simone Alessandri for his technical support. EA was supported by a grant from the Swiss National Science Foundation (13DPD6_130269) and by the Swiss State Secretariat for Education, Research, and Innovation. VJW was supported by a Vice-Chancellor's Fellowship from the University of Sheffield, UK.

References

[1] D. Hume, "Of the standard of taste," in *Four Dissertation*. London, 1757. Retrieved September 26, 2014 from <http://www.davidhume.org /texts/fd.html>

[2] H. Helmholtz, *Die Lehre von den Tonempfindungen, als physiologische Grundlage für die Theorie der Musik*, 4th ed. Braunschweig: Friedrich Vieweg und Sohn Verlag, 1877.

[3] J. H. McDermott, „Auditory preferences and aesthetics: Music, voices, and everyday sounds," in *Neuroscience of preference and choice: Cognitive and neural mechanisms*, R. J. Dolan & T. Sharot Eds. London: Academic Press, 2012, pp. 227–256.

[4] P. N. Juslin, S. Isaksson, "Subjective criteria for choice and aesthetic judgment of music: A comparison of psychology and music students," *Research Studies in Music Education*, vol. 36, pp. 179–198, 2014. doi: 10.1177/1321103X14540259

[5] S. Thompson, S. "Determinants of listeners' enjoyment of a performance," *Psychology of Music*, vol. 35, pp. 20–36, 2007.

[6] R. Reber, N. Schwarz, P. Winkielman, "Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience?" *Personality and Social Psychology Review*, vol. 8, pp. 364–382, 2004.

[7] M. J. Bergee, "Faculty interjudge reliability of music performance evaluation," *Journal of Research in Music Education*, vol. 51, pp. 137–150, 2003. doi: 10.2307/3345847

[8] D. W. Kinney, "Internal consistency of performance evaluations as a function of music expertise and excerpt familiarity," *Journal of Research in Music Education*, vol. 56, pp. 322–337, 2009. doi: 10.1177/0022429408328934

[9] W. J. Wrigley, S. B. Emmerson, "Ecological development and validation of a music performance rating scale for five instrument families," *Psychology of Music*, vol. 41, pp. 97–118, 2013. doi: 10.1177/03057 35611418552

[10] S. Thompson, A. Williamon, "Evaluating evaluation: musical performance assessment as a research tool," *Music Perception*, vol. 21, pp. 21–41, 2003.

[11] G. McPherson, E. Schubert, „Measuring performance enhancement in music," in *Musical Excellence*, A. Williamon, Ed. New York: Oxford University Press, 2004, pp. 61–82.

[12] C. Ryan, J. Wapnick, N. Lacaille, A. A. Darrow, "The effects of various physical characteristics of high-level performers on adjudicators' performance ratings," *Psychology of Music*, vol. 34, pp. 559–572, 2006.

[13] B. W. Vines, C. L. Krumhansl, M. M. Wanderley, D. J. Levitin, "Cross-modal interactions in the perception of musical performance," *Cognition*, vol. 101, pp. 80–113, 2006.

[14] M. Schutz, S. Lipscomb, „Hearing gestures, seeing music: Vision influences perceived tone duration," *Perception*, vol. 36, pp. 888–897, 2007.

[15] N. K. Griffiths, "Posh music should equal posh dress: an investigation into the concert dress and physical appearance of female soloists," *Psychology of Music*, vol. 38, pp. 159–177, 2010.

[16] M. Lehmann, R. Kopiez, „The influence of on-stage behavior on the subjective evaluation of rock guitar performances," *Musicae Scientiae* vol. 17, pp. 472–494, 2013. doi: 10.1177/1029864913493922
[17] F. Platz, R. Kopiez, „When the first impression counts: music performers, audience, and the evaluation of stage entrance behavior," *Musicae Scientiae*, vol. 17, pp. 167–197, 2013. doi: 10.1177/1029864913486369
[18] A. Gabrielsson, "Music performance research at the millennium," *Psychology of Music*, vol. 31, pp. 221–272, 2003.
[19] J. H. McDermott, „Auditory preferences and aesthetics: Music, voices, and everyday sounds," in *Neuroscience of preference and choice: Cognitive and neural mechanisms*, R. J. Dolan & T. Sharot, Eds. London: Academic Press, 2012, pp. 227–256.
[20] D. E. Lundy, "A test of consensus in aesthetic evaluation among professional critics of 900 modern music," *Empirical Studies of the Arts*, vol. 28, pp. 243–258, 2010.
[21] A. Van Venrooij, V. Schmutz, "The evaluation of popular music in the United States, Germany and The Netherlands: a comparison of the use of high art and popular aesthetic criteria," *Cultural Sociology*, vol. 4, pp. 395–421, 2010. doi: 10.1177/1749975510385444
[22] G. Guest, K. M. MacQueen, E. E. Namey, E. E. *Applied thematic analysis*. California: SAGE Publications, 2012.
[23] E. Alessandri, H. Eiholzer, A. Williamon, "Reviewing critical practice: An analysis of Gramophone's reviews of Beethoven's piano sonatas, 1923-2010," *Musicae Scientiae*, vol. 18, pp. 131–149, 2014. doi: 10.1177/1029864913519466
[24] D. J. Hopkins, G. King, "A method of automated nonparametric content analysis for social science," *American Journal of Political Science*, vol. 54, pp. 229–247, 2010.
[25] Y. R. Tausczik, J. W. Pennebaker, „The psychological meaning of words: LIWC and computerized text analysis methods," *Journal of Language and Social Psychology*, vol. 29, pp. 24–54, 2010.
[26] E. E. Namey, G. Guest, L. Thairu, L. Johnson, "Data reduction techniques for large qualitative data sets," in *Handbook for team-based qualitative research,* G. Guest & K. M. MacQueen, Eds. Maryland: Altamira Press, 2008, pp. 137–161.
[27] E. Alessandri, V. J. Williamson, H. Eiholzer, A. Williamon, "Beethoven recordings reviewed: a systematic method for mapping the content of music performance criticism," *Frontiers in Psychology*, vol. 6:57, 2015. doi:10.3389/fpsyg.2015.00057
[28] E. Alessandri, "The notion of expression in music criticism," in *Expressiveness in music performance: Empirical approaches across styles and cultures,* D. Fabian, R. Timmers & E. Schubert, Eds. Oxford: Oxford University Press, 2014, pp. 22–33.
[29] E. Alessandri, V. J. Williamson, H. Eiholzer, A. Williamon, "A critical ear: Analysis of value judgements in reviews of Beethoven's piano sonata recordings," *Frontiers in Psychology*, vol. 7:391, 2016. doi.org/10.3389/fpsyg.2016.00391
[30] V. J. Williamson, S. R. Jilka, J. Fry, S. Finkel, D. Müllensiefen, L. Stewart, "How do earworms start? Classifying the everyday circumstances of involuntary musical imagery," *Psychology of Music*, vol. 40, pp. 259–284, 2012. doi: 10.1177/0305735611418553
[31] V. J. Williamson, S. R. Jilka, S. R. "Experiencing earworms: an interview study of involuntary musical imagery," *Psychology of Music*, vol. 42, pp. 653–670, 2013. doi: 10.1177/0305735613483848
[32] N. Carroll, *On criticism*. New York: Routledge, 2009.
[33] M. C. Beardsley, "On the generality of critical reason," *Journal of Philosophy*, vol. 59, pp. 477–486, 1962.
[34] M. C. Beardsley, "The classification of critical reason," *Journal of Aesthetic Education*, vol. 2, pp. 55–63, 1968.
[35] S. Kaplan, R. Kaplan, *The Experience of Nature: a Psychological Perspective*. New York: Cambridge University Press, 1989.
[36] D. Davies, "Against enlightened empiricism," in *Contemporary debates in aesthetics and the philosophy of art*, M. Kieran, Ed. Oxford: Blackwell, 2006, pp. 22–36.
[37] G. Graham, "Aesthetic empiricism and the challenge of fakes and ready-mades," in *Contemporary Debates in Aesthetics and the Philosophy of Art*, M. Kieran, Ed. Oxford: Blackwell, 2006, pp. 11–21.
[38] G. Dickie, "Beardsley, Sibley and critical principles," *The Journal of Aesthetics and Art Criticism*, vol. 46, pp. 229-237, 1987.
[39] D. E. Berlyne, *Aesthetics and psychobiology*. New York: Appleton, 1971.
[40] R. Philip, *Performing music in the age of recording*. New Haven and London: Yale University Press, 2004.
[41] M. Katz, *Capturing sound: How technology has changed music*. Berkeley and Los Angeles: University of California Press, 2004.
[42] D. N. C. Patmore, E. F. Clarke, "Making and hearing virtual worlds: John Culshaw and the art of record production," *Musicae Scientiae*, vol. 11, pp. 269-293, 2007. doi: 10.1177/102986490701100206.
[43] S. Frith, "Going critical. Writing about recordings," in *The Cambridge Companion to Recorded Music*, N. Cook, E. F. Clarke, D. Leech-Wilkinson & J. Rink, Eds. New York: Cambridge University Press, 2009, pp. 267-282.
[44] S. Debenedetti, "The role of media critics in the cultural industries," *International Journal of Arts Management*, vol. 8, pp. 30-42, 2006. doi:10.2307/41064885
[45] A. Pollard, *Gramophone: The first 75 years*. Harrow, Middlesex: Gramophone Publications Limited, 1998.

Clouds and vectors in the spiral array as measures of tonal tension

Dorien Herremans[1] and Elaine Chew[1]
[1]Centre for Digital Music, Queen Mary University of London, UK

Musical tension is a complex composite concept that is not easy to define, nor quantify. It consists of many aspects, usually rooted in either the domain of psychology (e.g. expectation and emotion) or music (rhythm, timbre, and tonal perception). This research focuses on the latter category, more specifically, tonal tension. Four vector-based methods are proposed for quantifying and visualizing aspects of tonal tension based on the spiral array, a 3D model for tonality. The methods are evaluated empirically using variety of musical excerpts.

Our approach first segments a musical excerpt into equal length subdivisions and maps the notes to clouds of points in the spiral array. Using vector-based methods, four aspects of tonal tension are identified from these clouds. A first aspect, related to dissonance, is captured by the cloud diameter, which shows the dispersion of the note clusters in tonal space. The second aspect, related to tonal change, is the linear cloud momentum, which captures pitch set movements in the spiral array. Tensile strain measures how far the tonal center of each cloud is displaced from the global key. Finally, angular cloud momentum measures the directional change for movements in tonal space.

These methods are implemented in a system that visualizes their results as tension ribbons over the musical score, allowing for easy interpretation. The system was tested on stimuli used in Farbood (2012)'s empirical study. The results show that the user-annotated tension is almost always captured by one of the four methods, thus corroborating the effectiveness of our methods. The study reveals that tension is a composite characteristic that should not be captured by one fixed global measure, but by varying combinations of aspects such as those in this study. These results are verified thorough evaluation of snippets containing the Tristan Chord, excerpts from a Haydn string quartet, a Beethoven piano sonata, and Schoenfield's Trio and Cafe Music.

Narrative Experiences of Instrumental Pop Music

Elizabeth Hellmuth Margulis[1]
[1]Department of Music, University of Arkansas, Fayetteville, AR, USA

Despite colloquial accounts of hearing stories while listening to music, little empirical work has investigated this form of engagement. A series of studies in our lab examined the factors that contribute to narrative hearings of orchestral excerpts. Some people, however, tend to encounter orchestral music almost exclusively in the context of film. Therefore, we designed new studies that use instrumental excerpts from pop music as stimuli. People tend to hear this kind of music in a wider variety of circumstances, making it less exclusively reliant on cinematic associations. By comparing results, we can understand more about the relationship between enculturation and narrative listening.

50 participants heard 90s instrumental pop excerpts that either did or did not feature noticeable contrast (as identified by a preliminary study), pursuing the hypothesis that people resort to narrative to make sense of music featuring significant contrast (cf. Almen, 2008). After listening to each excerpt, participants completed the Narrative Engagement Scale for Music (NESM), adapted from Busselle & Bilandzic (2009), and responded to a number of other questions about their listening experience. At the end of the experiment, participants completed several measures including the Absorption in Music Scale (Sandstrom & Russo, 2013).

As in the studies that used classical excerpts, results showed a higher tendency to narrativize music that featured contrast. Also similar to previous studies, there was a surprising amount of convergence in free response descriptions of the narratives, with topical overlap that sometimes surpassed 80%. Contrastingly, however, the content of the free response descriptions provided in response to the pop excerpts differed in intriguing ways from those provided in response to the classical excerpts. The last part of the talk explores these differences and outlines an agenda for further work on the widespread behavior of narrative listening.

Personality Heard in Music

Erkki Huovinen

Department of Music, University of Jyväskylä, Finland / School of Music, North West University, South Africa
erkki.s.huovinen@gmail.com

ABSTRACT

Following composers' attempts to depict human characteristics in music, many musical theorists have suggested that music might be appreciated by the listener in terms of hearing human agency or "personae" in it. Despite the persistence of such ideas, there is a lack of empirical research in the personification of heard music. This study addresses two questions: (1) Can listeners' extramusical response to music reproduce the dimensions of human personality, as given by the Five-Factor Model, and (2) Is this dependent on listeners' cultural familiarity with the music? In preparation for the study, 38 Finnish music students selected their favorite pieces of instrumental music and judged each others' selections on 6 "mood variables" and 6 "human variables", helping to define a set of 10 excerpts meant to represent qualitative extremes in the participants' musical world. In Experiment 1, the same students first listened to these excerpts, imagining and writing down scripts for associated movie scenes involving a central protagonist. On another round of listenings, they were asked to carry out ("Short Five") personality tests concerning the imagined protagonists. A factor analysis of the results revealed five factors which nevertheless failed to match the traits of the Five-Factor Model. Instead, ingredients from several of the classical personality traits coalesced into factors called "goodness," "heroism," and "weakness." In Experiment 2, the same procedure, with the same music, was repeated with 37 South African music students. Somewhat surprisingly, the listeners' judgments now quite exactly reproduced the traits of the Five-Factor Model. To conclude, associative features related to human personality seem to be widely available as tools for the appreciation of even unfamiliar music, but cultural knowledge may further bias such associations in terms of culturally established stereotypes.

I. INTRODUCTION

Most of the empirical research concerning the relationship between music and personality has focused on how dimensions of personality might be associated with human beings' reactions to, likings of, and activities around music. Much of the early literature on the topic was centered on musicians' personality profiles, including findings regarding, say, musicians' tendency toward introversion (for a review, see Kemp 1996). Later, there has been more emphasis on the relationships between personality, on the one hand, and musical listening habits and everyday musical activities and phenomena, on the other. For example, listeners' personality dimensions have been shown to be associated with their musical preferences (Rentfrow & Gosling 2003), their emotional reactions to music (Ladinig & Schellenberg 2012), and their ways of using music in everyday life (Chamorro-Premuzic & Furnham 2007).

There is also another, longer intellectual tradition that finds a different link between personality and music. Here, the question is not about how people with different personalities might shape their musical lives differently, but rather about whether music, as heard sound, might also come to be heard as expressive of human personality. This idea is a commonplace in the history of Western classical music, where we often find composers depicting human persons in music: one may come to think of such examples as romantic tone poems named after historical or legendary figures (*e.g.*, Strauss' *Don Juan*), or Wagner's use of leitmotifs to characterize dramatic figures in his works. Composers have also been interested in the concept of personality as such, dedicating whole compositions to the four classical temperaments (*e.g.*, C. P. E. Bach's Trio Sonata H. 579, "Dialogue between a *Melancholicus* and a *Sanguineus*," Carl Nielsen's Symphony No. 2, Op. 16, "The Four Temperaments," and Paul Hindemith's *Theme and Four Variations* (*The Four Temperaments*)). In its references to personality, musicological thinking, too, has often relied on ideas inspired by this expressive tradition. Hence Spitzer (2013, 11) notes that "Schubert's stand-alone songs comprise a veritable *comedie humaine* of personality types" and Kramer (1990, 5) states that "music does have referential power, even if we are not prepared to be very precise about it. To affirm, for example, that nineteenth-century overtures named for Coriolanus, Manfred, and Hamlet fail to represent those characters seems foolish if not perverse."

In musical scholarship, such ideas have also spawned more specific theories concerning the personification of music—theories of a kind that have remained conspicuously aloof from empirical music psychology. Two types of claims could be usefully distinguished here. First, there are the semiotic approaches offering critical interpretations of musical works in terms of actoriality or agency. Tarasti connects musical actoriality with distinctive themes or motifs: for him, it is these musical units themselves that consitute actors, or "musical subjects that influence and perform within the musical discourse" (Tarasti 1994, 122). Hatten extends Tarasti's scheme by noting that besides musical agents on the "level of story" (often in the opposing roles of protagonists and antagonists), these may be fused into one principal agent, or "narrative persona which, in telling its own tale, can also direct its own adventures" (*ibid.*, 230). It should be noted that these theorists do not principally discuss personification as it might be immediately relevant to everyday listeners' understanding of music: in musical semiotics, the main questions are about interpretative choices justified by critical reasoning. Hatten, for instance, speaks of difficulties in *deciding* among various kinds of agency, and of *weighting* them in various ways in critical interpretations of musical scores (*ibid.*, 225, 230). For present purposes, such interpretive lines of music scholarship nevertheless indicate a possibility that seeds for personifying interpretations might be found even more generally in musical listening. Hatten, indeed, implies this by suggesting that as soon as music is heard in gestural terms, there is also an implication of agency, because a musical gesture "communicates information about the gesturer" (Hatten 2004, 224).

Apart from semiotic theories aiming at detailed, interpretative analyses of particular musical works, some scholars have also presented full-scale *persona theories* which go a step further toward generalized claims. Foreshadowing the above-mentioned distinction between story-level agents and an overarching narrative persona, Edward T. Cone (1974) made a distinction between vocal or instrumental protagonists, on the one hand, and what he called the *complete musical persona*, on the other. For Cone, "any instrumental composition […] can be interpreted as the symbolic utterance of a virtual persona. […] [I]n every case there is a musical personality that is the experiencing subject of the entire composition, in whose thought the play, or narrative, or reverie, takes place—whose inner life the music communicates by means of symbolic gesture" (Cone 1974, 94). The complete musical persona is "by no means identical with the composer; it is a projection of his musical intelligence, constituting the mind, so to speak, of the composition in question" (*ibid.*, 57). In discussing Cone's theory, Robinson (2005, 329), suggests that it may help the critic "explain what appear to be anomalies in a piece when it is interpreted as a 'purely musical' structure of tones." Again, the question would not be whether and how impersonation actually figures in uninformed musical listening, but rather whether and how one might find justification for critical practices drawing on personification. In Robinson's (2005, 332) view, "it is appropriate to interpret a piece of instrumental music as containing characters if this interpretation is consistent with what is known about the composer, including his compositional practices and beliefs, and if interpreting the piece in this way is part of a consistent interpretation that makes sense out of the piece as a whole."

The philosopher Jerrold Levinson has gone further, however, in proposing a roughly similar theory as a general *definition* of musical expressiveness. Briefly, he takes a piece of music to be expressive of a given psychological condition α *if and only if* the piece is "most readily and aptly heard by the appropriate reference class of listeners as (or as if it were) a sui generis personal expression of α by some (imaginatively indeterminate) individual" (Levinson 1990, 338). One problem with such a formulation is the logical equivalence relation ("if and only if") that suggests musical expressivity to be present in music *in all and only those situations* in which the appropriate reference class of listeners will tend to hear the music as an expression of an imagined persona. This may be too strong a claim (for a discussion, see Huovinen & Pontara 2011). For present purposes, what is more important is the mere fact that Cone, Levinson, and Hatten would all suggest a way of hearing music as a temporally extended experience of or expression by a persona in the music, fluctuating in its nuances as the music changes and progresses. We should ask: do other listeners, too, more generally hear music in such terms?

In discussing this issue with music students, I have found that even despite their interest in conceptualizing music-related phenomena, they often find it hard to accept persona theories as intuitively credible, or even as intelligible. This could be because they are not as musically sensitive as our theorist troika, or not sophisticated enough in their critical acumen—or because the persona theorists themselves are completely misguided. Such explanations would leave persona theories the sole province of critical scholasticism. Another explanation might be that the persona theorists, in hypostatizing entities that might seem foreign to many listeners' everyday musical understanding, have been unable to communicate the idea of music's impersonation in an accessible enough manner. In this paper, I suppose that this is indeed the case, and I want to make the suggestion that the source of persona theories lies simply in the ease with which listeners may associate music with various significant spheres and aspects of human life, and hence, potentially, with significant dimensions of the human personality as we know it. In other words, to say that a piece of music is easily heard as an expressive utterance of depression by a virtual persona is perhaps just a way of saying that *when* listeners' mindsets or contexts, for one reason or another, suggest associating the music with thoughts about oneself or a fellow human being, the complex associations awoken (in the reference listener population) tend to be toward the depressive side of human existence. From this angle, claims concerning the impersonation of music would not hinge on positing virtual personae or other mysterious entities "in the music." Rather, the core claim would simply be that in a given cultural setting, a call to associate human personality characteristics with a range of musical genres, styles, and textures may produce a roughly consistent mapping. In the following study, this claim will be evaluated by assuming the received Five-Factor Model (Costa & McCrae 1992; see, *e.g.*, Matthews et al. 2003) as a conveniently fine-grained model of human personality, and by using the participant group's own favorite musics as a basis for sampling the range of culturally relevant musical sounds.

This study addresses two questions: (1) Can listeners' extramusical response to music reproduce the dimensions of human personality, as given by the Five-Factor Model, and (2) Is this dependent on listeners' cultural familiarity with the music? Basically, then, we want to know whether heard music is personified in listeners' understanding in ways that convey the "Big Five" traits of Neuroticism (N), Extraversion (E), Openness (O), Agreeableness (A), and Conscientiousness (C).

These questions are approached by finding the "qualitative extremes" of the favorite instrumental music of a listener group in one culture, and by subjecting this music to a kind of "personality analysis" by this group of listeners as well as by another group of listeners in a different culture. The basic idea is simple: on a first round of listenings, the participants would imagine and write down scripts for associated movie scenes involving a central protagonist. On another round of listenings, they would then be asked to fill out standard personality questionnaires concerning the imagined protagonists. Finally, factor-analytical solutions of the questionnaire results would, for both of the listener groups, indicate the extent to which listeners' imagery of human characters in response to the music might reproduce something like the dimensions of the Five-Factor Model. The working hypothesis was that this would indeed happen in the "home culture" from which a qualitatively varied sample of music is culled, but that such reproduction of the Five-Factor structure might not succeed with a group of listeners less culturally familiar with the music. If this is right, the "personality dimensions" emerging from the confrontation of listeners with culturally foreign music might be either relatively coarse (with fewer than five factors emerging in the analysis), or they might be differently organized (reflecting a less clear understanding of the music as "naturally" expressive of human personality).

II. EXPERIMENT 1

With reference to the overall research plan presented above, Experiment 1 was intended to test the hypothesis that listeners would tend to associate music from their own musical culture with human characteristics in such a manner as to roughly reproduce the dimensions of the Five-Factor Model of personality. Accordingly, the Finnish listeners taking part in Experiment 1 would first be asked to select their favorite instrumental music to be used in the study. In the experiment, they would hear music representing qualitative extremes of musical genres known and valued by themselves as a group—by young Finnish adults with an interest in music.

A. Method

1) Participants. The participants were 38 Finnish university students of musicology (incl. 28 majors), with a mean age of 25.4 years (sd = 7.7). They had played or sung music actively for an average of 14.5 years (sd = 7.8), and had participated in formal music education for an average of 9.4 years (sd = 5.7). 18 of them reported playing or singing music daily, and the rest at least once a week. Most of them indicated having had some experience of improvising (33) and/or composing (21) music during the past year. The participants also reported averages of 16.5 hours of weekly listening to music (sd = 15.9), 7.6 hours of playing or singing music (sd = 6.6), and 6.2 hours of watching movies or TV (sd = 6.5).

2) *Preparatory Phase: Music Sampling.* The purpose of this preparatory phase was to sample the qualitative extremes of the participants' musical environment by having them evaluate each others' favorite pieces of music with regard to their mood and their closeness to existential aspects of human life. Each participant was given the task to select one recorded piece of instrumental music (in any genre, style, or tradition) which was personally important for him/her, and to upload it as a soundfile on a course website. With two students selecting the same piece, and one self-recorded excerpt discarded, the resulting pool of musical selections comprised 36 pieces that, as a whole, gave an impression of the musical tastes among the target group of listeners. In all, there were 16 selections of classical music, 8 pieces representing various genres of rock music, 5 pieces of film or TV music, 3 musical excerpts from video games, as well as some jazz (2), folk (1), and gospel (1) titles.

These pieces were edited to briefer versions of a 1'30''–1'35'' duration (with a 3 s linear fade-out), starting either at the beginning of the original recording, or, in cases where the participant had suggested a more specific favorite location, starting from there (with a 3 s fade-in, if needed). The edited excerpts were arranged in random order and uploaded again on the course website, where the participants could individually listen to them at their leisure. To counterbalance effects of tedium, half of the participants saw the list of musical excerpts in the opposite order. After listening to each piece, each participant wrote down the name and composer or performer of the excerpt if s/he knew it, and judged the "fittingness of given opposite words to describe the mood of the music." The 12 bipolar, 7-point rating scales (between –3 and 3) were conceived in terms of six adjectival *mood variables* (three dynamic ones and three sensory ones) as well as six nominal *human variables* (three with teleological and three with affirmative import; see Table 1 for English translations of the original Finnish terms). The task was carried out in the time frame of 4–14 days before the main experiment, thus familiarizing the participants with the music beforehand.

Table 1. The mood and human variables used in the preparatory phase of Experiment 1

mood variables		human variables	
dynamic	sensory	teleological	affirmative
moving/static	warm/cold	future/past	reality/imagination
tense/relaxed	light/dark	desire/contentment	life/death
strong/weak	soft/hard	fate/chance	knowledge/doubt

The participants' overall familiarity with each others' favorite music was not very high: on average, they knew only 6.1 of the other participants' music selections by either the name and the composer/performer (sd = 3.7), and only 3.8 by both (sd = 3.0).

Out of the 36 musical excerpts evaluated, a set of ten pieces was selected to be used in the main experiment. The selection process aimed at sampling the qualitative extremes of this pool of music, by first arranging the 36 pieces in 12 ranking orders according to the mean results on each of the scales in Table 1. Those pieces that most often appeared at rank orders 1–3, counted from either end of the ranking, were taken as the core of the stimulus set. Hence, as can be gleaned from Table 2, the dramatic beginning of Shostakovich' *Fifth Symphony* would appear at these rank orders for no less than seven of the scales: in relation to the other music heard during Phase 1, this music was experienced as extremely tense, strong, cold, and dark, as well as reflecting desire, death, and doubt. Conversely, the 1972 rock instrumental *Sylvia* by the Dutch group Focus would be heard as light (vs. dark), and full of contentment, reality, life, and knowledge. Other pieces with such extremely clear profiles included the beginning of Prokofiev's *Piano Sonata No. 7* (moving, tense, cold, hard, chance, doubt), a piece by the French post-metal band Year of No Light (tense, strong), and jazz pianist Bill Evans' signature piece *Waltz for Debby* (relaxed, warm, soft). As the mean rankings shown in Table 2 reveal, these pieces tended to appear near the extremes for most of the scales. In order to maximize variety, additional selections were made from those pieces that had extreme values on one or two of the scales and did not match any of the already chosen pieces in their qualitative profiles. Hence the opening of Alfredo Casella's harp sonata was experienced as congenial to imagination, the

Table 2. The 10 pieces of music eliciting the strongest responses, and their rankings on the mood and human variables (the top and bottom three rank orders have been bolded)

	mood variables						human variables						
	moving	tense	strong	warm	light	soft	future	desire	fate	reality	life	knowledge	mean ranking[1]
CLASSICAL MUSIC													
Casella, *Sonata per arpa*	6	18	29	16	17	16	11	25	32	**36**	14	29	11.00
Prokofiev, *Piano Sonata No. 7*	**1**	**1**	5	**35**	30	**35**	4	10	**34**	32	22	**35**	4.75
Satie, *Gymnopédie 1*	**36**	**35**	33	10	13	5	25	26	23	21	15	19	10.08
Shostakovich, *Symphony No. 5*	24	**3**	**3**	**36**	**36**	33	26	**3**	5	8	**35**	**36**	4.58
Sibelius, *Violin Concerto*	22	5	12	31	24	27	17	**1**	13	10	31	33	9.33
ROCK MUSIC													
Focus, *Sylvia*	5	28	9	7	**2**	17	5	**36**	27	**2**	**1**	**1**	5.75
Gary Moore, *The Loner*	18	17	7	21	21	24	**35**	11	**2**	**1**	22	11	10.75
Year of no light, *Persephone*	8	**1**	**1**	32	33	32	7	6	10	7	28	9	6.00
JAZZ MUSIC													
Bill Evans, *Waltz for Debby*	12	**36**	30	**1**	9	**3**	8	32	33	17	5	7	6.58
GAME MUSIC													
Uematsu, *Interrupted by Fireworks*	23	32	**36**	13	**3**	10	9	30	22	**34**	12	13	8.75

[1] Means for ranking numbers taken from the closer end of the scale (36 = 1, 35 = 2, etc.).

opening of Sibelius' violin concerto sounded of desire, Erik Satie's simplistic piano piece *Gymnopédie 1* was static and relaxed, Max Middleton's rock ballad *The Loner* in guitarist Gary Moore's 1987 rendition resonated with the past, fate, and reality, and Nobue Uematsu's music for the 1997 video game *Final Fantasy VII* was experienced as weak and light, and relevant for imagination.

The set of music sampled thus not only evoked strong immediate judgments in the participant group, but also did this in sharply distinct fashions. It may be noted that none of the film and TV music—supposedly engineered to please as large groups of listeners as possible—or the pieces most familiar to the participant group (Tchaikovsky's *Swan Lake* and Beethoven's *Moonlight Sonata*) made it to the selection based on strong responses. Indeed, prior familiarity did not seem to play much of a role in eliciting strong responses: Shostakovich' Fifth Symphony, for instance, was only recognized by three of the participants, and hence the darkness, doubt and death heard in it could not stem from associations to the composer's harsh conditions under the Soviet rule.

3) *Procedure, Phase 1: Imagined Scenes.* In the first phase of the main experiment, the participants listened to the 12 musical excerpts in a quasi-random order in a classroom setting. On listening to each excerpt, their task was to imagine and describe a human character that the music could suitably depict. So as to avoid alienating persona constructions, the task was framed with reference to choices standardly made in film production—that is, in cultural practices the participants would be aware of and probably esteem. The following instructions were given before the experiment:

> You will here ten musical excerpts. With each of these pieces of music, imagine that you are a film director who has carefully selected music for a scene depicting a character in the film. Listen to the music with your eyes closed, trying to be sensitive to the mood of the music, and imagine what sort of a character and scene could be in question. Do not ponder too much, but try to grasp your first intuition!
>
> When you have arrived at a suitable character and a scene, open your eyes and write a description addressing at least (1) who the person depicted in the scene is, (2) what he or she is like as a person, and (3) what is happening in the scene.
>
> Try to write freely and spontaneously, without unnecessary self-criticism. The more descriptive and detailed your writing, the easier it will be to carry out Part 2 of this experiment. Therefore, the richness of your descriptions is more important than linguistic correctness or textual structure.

4) *Procedure, Phase 2: Personality Analysis.* After completing Phase 1, and after a short break, Phase 2 of the main experiment involved listening to the pieces of music again and filling out standard personality questionnaires for each of the previously imagined protagonists. The instructions were as follows:

> You will now hear the previous musical excerpts once again. Imagine that for each piece of music, your task as film director is to describe to the actor or actress what sort of character he or she would be portraying in the scene underscored by the music.

> With each musical excerpt, listen to the piece, read your previous description about the film character, and try to connect with the mood of the scene you described. […] When the music comes to a close, analyze the personality of the film character by assessing him or her with the given adjectives. […]
>
> While responding, try to keep the imagined character constantly in your mind and empathize with his/her life situation in order to describe the person as accurately as possible.

The personality questionnaire, administered separately for each of the imagined characters, was the "Short Five" (S5) that measures the 30 facets of the Five-Factor Model with a set of 60 comprehensive single items—that is, pairs of positive and negative indicators intended to match expert descriptions of the constructs as closely as possible. For the Finnish and English versions of S5 (used in the two experiments of the present study), Konstabel et al. (2012) report factor level correlations ranging from .78 (English A) to .91 (Finnish E) with the longer NEO PI-R personality inventory (in the same language), as well as a factor structure highly similar to the normative US NEO-PI-R sample (with all congruence coefficients exceeding .89).

B. Results

The personality analyses from the second phase of the experiment were subjected to exploratory factor analysis in the software *R*. Parallel analysis suggested five factors to be extracted. The factor analysis, applying the minimum residual factoring method with promax rotation, produced the pattern matrix shown in Table 3 (loadings of at least .40 are shown). For ease of reading, the table is organized in five panels, corresponding to the dimensions of the Five-Factor Model of personality—Neuroticism (N), Extraversion (E), Openness (O), Agreeableness (A), and Conscientiousness (C).

Unlike expected, the extracted five-factor structure appreciably deviated from the patterns of covariation suggested by the Five-Factor Model. In fact, only F4 faithfully reproduced the Conscientiousness dimension as such, whereas the other factors have to be given tentative interpretations that do not match the "Big Five." Factor F1, although encompassing the whole trait of Agreeableness, also covers such items as Openness to values (O6), Openness to feelings (O3), Warmth (E1), and *no* Angry hostility (N2). This factor thus appears as some kind of general *Goodness* dimension. Even more interestingly, F2 quite evenly mixes concepts from four of the Big Five dimensions—Excitement seeking (E5), Openness to actions (O4), Activity (E4), Impulsiveness (N5), Openness to ideas (O5), Achievement striving (C4), and Competence (C1). I would like to call this a *Heroism* dimension: it appears to delineate a character not unlike the generic image of a movie action hero. F3, in turn, covers much of Neuroticism, but also the *lack* of the constructs Positive emotions (E6), Assertiveness (E3), and Trust (A1), perhaps suggesting overall *Weakness*. Finally, F5, as the least significant of the factors, simply appears as a reduced version of A, suggesting something like *Honesty*.

Table 3. Imagined persons (Finland): Factor loadings and communalities for the 30 items of the Short Five questionnaire.

	F1	F2	F3	F4	F5	com.
Anxiety (N1)			.92			.80
Angry hostility (N2)	-.51		.49			.60
Depression (N3)			.82			.68
Self-consciousness (N4)			.65			.46
Impulsiveness (N5)		.76				.49
Vulnerability (N6)			.80			.55
Warmth (E1)	.74					.78
Gregariousness (E2)	.45					.62
Assertiveness (E3)			-.55			.66
Activity (E4)		.80				.67
Excitement seeking (E5)		.84				.72
Positive emotions (E6)			-.56			.79
Openness to fantasy (O1)						.45
Openness to aesthetics (O2)	.44					.47
Openness to feelings (O3)	.66					.57
Openness to actions (O4)		.81				.72
Openness to ideas (O5)		.57				.60
Openness to values (O6)	.74					.68
Trust (A1)	.43		-.45			.62
Straightforwardness (A2)	.88				.41	.64
Altruism (A3)	.99				.53	.81
Compliance (A4)	.89					.82
Modesty (A5)	.80					.57
Tender-mindedness (A6)	.94					.78
Competence (C1)		.41		.49		.60
Order (C2)				.88		.74
Dutifulness (C3)	.42			.60		.57
Achievement striving (C4)		.44		.45		.51
Self-discipline (C5)				.83		.67
Deliberation (C6)				.80		.66
Eigenvalues	8.74	5.09	3.11	1.67	0.71	
% of variance	.23	.14	.14	.10	.02	
Cumulative variance	.23	.37	.52	.62	.64	

C. Discussion of Experiment 1

Imagining heard music to accompany film scenes and analyzing the personalities of protagonists inhabiting these scenes provided an intuitively acceptable way of probing the relevance of personification theories. Judging by the way in which participants began writing down their scripts rather quickly and effortlessly, the task seemed to have a degree of ecological validity for them: instead of being persuaded to hear the music as the personal expression of a persona residing in the music, they could relate the task to the familiar issue of how music accompanies narrative in movies.

Before carrying out the study, it was hypothesized that such a task, in conjunction with a broad enough range of more or less familiar musical sounds and styles, could subsist a "natural" categorization of human personality dimensions. Instead of reproducing the personality dimensions suggested by the received Five-Factor Model, however, the five factors extracted from personality questionnaire data concerning the imagined protagonists showed a very different structure. In particular, the first three factors—tentatively labeled Goodness, Heroism, and Weakness—grouped the questionnaire items in ways departing from the standard Big Five conception. An obvious reason suggests itself: the participants were, in general, heavy media users, and hence the range of movie protagonists imagined to accompany the heard musical examples might also reflect the participants' knowledge of film and popular culture rather than just their knowledge of real-world persons. In sum, the participants quite consistently converged on distinct "personality profiles" for the musical excerpts, but it seems that these profiles were organized less by real human psychology and more by stereotypical narrative figures ingrained by popular culture.

III. EXPERIMENT 2

In Experiment 1, the main result (that human imagery catalyzed by the music failed to reproduce the dimensions of the Five-Factor Model of personality) was obtained by means of music with relative cultural familiarity to the participants. Even though the particular pieces of music were not very often recognized by the listeners, the stylistic and genre characteristics of the music may nevertheless have been quite familiar to them. Without some cross-cultural comparison, it would be hard to assess the extent to which cultural knowledge may have played a role in shaping the patterns of covariation seen in the factor analysis of the results. The purpose of Experiment 2 was, therefore, to take the musical materials selected in Finland and repeat the main two phases of the experiment with a comparable group of listeners from a very different culture. South Africa was chosen because its relative socio-cultural and music-cultural distance to Northern Europe.

A. Method

1) Participants. The participants were 37 South African university music majors (incl. 14 voice majors), with a mean age of 21.5 years (sd = 2.7). They had played or sung music actively for an average of 10.3 years (sd = 4.8), and had participated in formal music education for an average of 4.3 years (sd = 4.2). 24 of them reported playing or singing music daily, and the rest at least once a month. Most of them indicated having had some experience of improvising (34) and/or composing (30) music during the past year. The participants also reported averages of 25.1 hours of weekly listening to music (sd = 21.7), 14.6 hours of playing or singing music (sd = 12.3), and 5.9 hours of watching movies or TV (sd = 6.6).

2) Musical materials and procedure. The experiment was conducted in a classroom setting, with the same musical materials as in Experiment 1, and repeating the same succession of two phases (Phase 1: Imagined Scenes, Phase 2: Personality Analysis).

B. Results

The personality questionnaire data from the second phase of the experiment was subjected to a similar exploratory factor analysis as in Experiment 1. Parallel analysis suggested five factors to be extracted. A factoring solution with promax rotation produced the pattern matrix shown in Table 4 (loadings of at least .40 are shown). As should immediately be clear, the South African participants have analyzed their imagined protagonists in a manner that neatly reproduces the five personality dimensions of the Five-Factor Model.

Table 4. Imagined persons (South Africa): Factor loadings and communalities for the 30 items of the Short Five questionnaire.

	F1	F2	F3	F4	F5	com.
Anxiety (N1)			.80			.57
Angry hostility (N2)			.81			.62
Depression (N3)			.61			.55
Self-consciousness (N4)			.77			.59
Impulsiveness (N5)			.67			.38
Vulnerability (N6)			.64			.53
Warmth (E1)	.45					.52
Gregariousness (E2)				.40		.58
Assertiveness (E3)				.54		.55
Activity (E4)				.89		.69
Excitement seeking (E5)				.88		.77
Positive emotions (E6)				.57		.62
Openness to fantasy (O1)					.70	.48
Openness to aesthetics (O2)					.62	.48
Openness to feelings (O3)					.63	.46
Openness to actions (O4)					.62	.53
Openness to ideas (O5)					.72	.44
Openness to values (O6)					.52	.52
Trust (A1)	.53					.55
Straightforwardness (A2)	.62					.58
Altruism (A3)	.80					.58
Compliance (A4)	.96					.75
Modesty (A5)	.84					.63
Tender-mindedness (A6)	.72					.63
Competence (C1)		.50				.43
Order (C2)		.83				.70
Dutifulness (C3)		.73				.55
Achievement striving (C4)		.63				.61
Self-discipline (C5)		.81				.67
Deliberation (C6)		.80				.58
Eigenvalues	9.67	3.47	1.70	1.45	0.85	
% of variance	.14	.12	.11	.10	.09	
Cumulative variance	.14	.26	.37	.48	.57	

C. Discussion of Experiment 2

Unlike initially expected, increasing the cultural distance between the listeners and the heard music did not reduce the dimensionality of personality associations induced by the music, nor did it push the patterns of covariation in these associations farther away from what would be likely in standard applications of personality inventories. Instead, the number of distinct factors remained the same in the South African and Finnish cases and, more importantly, the dimensional structure of personality judgments for the imagined protagonists actually *approached* the standard human pattern in the case of culturally more distant listeners. It seems, then, that despite any cultural "irrelevance" that might be claimed for using such "foreign" music in a music-psychological experiment with South African participants, these listeners were, in fact, highly capable of grasping the music as expressive of important dimensions of human existence. While they may have been lacking in some of the implicit cultural knowledge concerning typical extramusical links for the relevant musical sounds and genres, the South African listeners approached the spectrum of heard musical sounds with a spectrum of more general human characteristics in mind.

IV. GENERAL DISCUSSION

In audiovisual media, music is often used to illustrate distinctions in human psychology. In classical Hollywood movies, character types were conspicuously connected with musical signifiers (Gorbman 1987), but there are also subtler possibilities for using music in cinema to convey a sense of represented subjectivity: music may sometimes offer a glimpse, as it were, to the film character's inner world (Stillwell 2007; Pontara 2014). In these and other ways, our omnipresent media cultures create and strengthen associations between music and human personality. This is but one aspect of a broader phenomenon of musical topicality, whereby sets of musical pieces, styles, and genres are gradually furnished with extramusical meaning in populations of listeners (see Huovinen & Kaila 2015). Such cultural processes of collective meaning-making take place on a vast time scale, making it hard to draw distinctions between culturally contingent associations and ones that might be more deeply rooted in, say, analogical gestural patterns between the music and the extramusical world. This article has been an attempt to address such distinctions—not by searching for universally recognizable "musemes" (small units of musical meaning; see, *e.g.*, Tagg 2013), but rather by focusing on music-driven associations with complex, high-level psychological properties, and by examining the internal structuredness of such associations in cases of smaller and greater cultural distance between the music and the listeners.

In this study, Finnish and South African listeners listened to pieces of music representing some of the favorite music of the Finnish group, and were asked to imagine film characters that the music might suitably depict. By having the participants fill out personality questionnaires for their imagined characters, we could employ the standard factor-analytical methods of personality research to extract the dimensions of judgments involved in the listeners' human imagery. The methodological basis for the study was given by the Five-Factor Model of personality, which has been found to have cross-cultural validity as a model of human personality trait structure (McCrae & Costa 1997). The result that our South African listeners' "personality analysis" of the music-inspired protagonists closely matched the trait structure of the Five-Factor Model suggests that the same dimensional structure that cross-culturally emerges from people's judgments concerning their own personality can also emerge from their associative imagery catalyzed by music—even by music with some cultural distance from their own cultural setting. This could be viewed as another way of putting forth the old idea of music as a "universal language": different expressive varieties of music may cross-culturally function as roughly reliable markers of universally important dimensions of human personality.

Even though associative features related to human personality may thus be cross-culturally available as tools for the appreciation of unfamiliar music, specific cultural knowledge may further bias such associations towards culturally established stereotypes. In the present study, the Finnish group of listeners, in their associative imagery evoked by "their own music," did *not* conform to the dimensional structure suggested by the Five-Factor Model. The whole pattern of results was, then, contrary to the working hypothesis behind the study. It had been originally supposed that more specific cultural knowledge of the musical genres and stylistic features might allow listeners to better appreciate a range of such genres and styles as bearers of information concerning the fine distinctions in human personality. Even if this were true to some small extent, any such increase in acuity was here clearly superseded by the cultural insiders' knowledge of the topical associations between varieties of music and varieties of human personality. For these listeners, the narrative imagery evoked by the music did not so much revolve around "normal" human characters as it did around stereotypical figures ingrained by popular culture such as the generic "good person" or the "action hero." In sum, it seems that music is a medium with potential universal significance as a signifier of human personality characteristics, but such significance can be overridden by culturally constructed and instilled understandings of extramusical meaning in music.

With regard to the persona theories discussed in the introduction, these findings suggest that the personification of music does not need to be treated solely as a question of careful critical interpretation of musical works. The participants in the present study were admittedly music students with perhaps a keener interest in music's meanings than one would find in a general population, but they did not have access to musical scores or information about the genesis of the pieces heard, and they could not approach the fleeting experimental tasks with the kind of rational deliberation and consideration of alternative interpretive possibilities that is characteristic of semiotic persona theorizing. Instead, they improvised their imagined scenes on the fly while listening, grasping the music in more intuitive terms, and nevertheless appeared to describe a world as rich in its dimensions of personality information as the real human world is.

In assessing such results, there is no reason to claim, *à la* Levinson, that musical expressivity *consists* in impersonation, or should be *defined* in terms of it. It is enough to say that at will, listeners across cultures, in response to heard music, can be expected to be capable of easily directing their thoughts in ways that delineate either more universal or, perhaps, more culturally impregnated conceptions of human personality. As soon as the listener, for one reason or another, activates person-related thoughts or emotions in the listening situation, the music will be readily available to shape, sharpen, focus, or rechannel this human imagery. No hypothetical musical personae are needed to make this happen. It is just that our lives as human persons, in societies made up of other human persons, are so imbued with person-related thinking that personality-relevant associative modes are always awaiting to be activated when heard music gets our minds to wander into our memories, fears, hopes, plans, and dreams.

ACKNOWLEDGMENT

This research was supported by a Research fellowship in the Research niche area Musical Arts in South Africa: Resources and Applications (MASARA) at the North West University, Potchefstroom campus, South Africa. Thanks are also due to prof. Jan-Erik Lönnqvist at the University of Helsinki for advice in using the Short Five personality inventory.

REFERENCES

Chamorro-Premuzic, T. & Furnham, A. (2007). Personality and music: Can traits explain how people use music in everyday life? *British Journal of Psychology*, *98*(2), 175–185.

Cone, E. (1974). *The composer's voice*. Berkeley, Los Angeles, and London: University of California Press.

Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources, Inc.

Gorbman, C. (1987). *Unheard melodies: Narrative film music*. Bloomington & Indianapolis, IN: Indiana University Press.

Hatten, R. S. (2004). *Interpreting musical gestures, topics, and tropes: Mozart, Beethoven, Schubert*. Bloomington and Indianapolis: Indiana University Press.

Huovinen, E. & Kaila, A.-K. (2015). The semantics of musical topoi: An empirical approach. *Music Perception*, *33*(2), 217–243.

Huovinen, E. & Pontara, T. (2011). Methodology in aesthetics: The case of musical expressivity. *Philosophical Studies*, *155*(1), 45–64.

Kemp, A. E. (1996). *The musical temperament: Psychology and personality of musicians*. New York: Oxford University Press.

Konstabel, K., Lönnqvist, J.-E., Walkowitz, G., Konstabel, K., & Verkasalo, M. (2012). The "short five" (S5): Measuring personality traits using comprehensive single items. *European Journal of Personality*, *26*(1), 13–29.

Kramer, L. (1990). *Music as cultural practice, 1800–1900*. Berkeley, Los Angeles, and Oxford: University of California Press.

Ladinig, O. & Schellenberg, E. G. (2012). Liking unfamiliar music: Effects of felt emotion and individual differences. *Psychology of Aesthetics, Creativity, and the Arts*, *6*(2), 146–154.

Levinson, J. (1990). *Music, art, and metaphysics: Essays in philosophical aesthetics*. Ithaca and London: Cornell University Press.

Matthews, G., Deary, I. J., & Whiteman, M. C. (2003). *Personality traits*. 3rd ed. Cambridge and New York: Cambridge University Press.

McCrae, R. R. & Costa, P. T., Jr. (1997). Personality trait structure as a human universal. *American Psychologist*, *52*(5), 509–516.

Pontara, T. (2014). Bach at the space station: Hermeneutic pliability and multiplying gaps in Anderi Tarkovsky's *Solaris*. *Music, Sound, and the Moving Image*, *8*(1), 1–23.

Rentfrow, P. J. & Gosling, S. D. (2003). The do re mi's of everyday life: The structure and personality correlates of music preferences. *Journal of Personality and Social Psychology*, *84*(6), 1236–1256.

Robinson, J. (2005). *Deeper than reason: Emotion and its role in literature, music, and art*. New York: Oxford University Press.

Spitzer, M. (2013). Sad flowers: Analyzing affective trajectory in Schubert's "Trockne Blumen." In T. Cochrane, B. Fantini, & K. R. Scherer (Eds.), *The emotional power of music: Multidisciplinary perspectives on musical arousal, expression, and social control* (pp. 7–21). Oxford: Oxford University Press.

Stillwell, R. (2007). The fantastical gap between the diegetic and nondiegetic. In D. Goldmark, L. Kramer, & R. Leppert (Eds.), *Beyond the soundtrack: Representing music in cinema* (pp. 184–202). Berkeley & Los Angeles: University of California Press.

Tagg, P. (2013). *Music's meanings: A modern musicology for non-musos*. New York & Huddersfield: The Mass Media Music Scholar's Press, Inc.

Tarasti, E. (1994). *A theory of musical semiotics*. Bloomington and Indianapolis: Indiana University Press.

Effects of Range and Repetition on Liking for Melodies

David Temperley[1], Celeste Kidd[2], and Amanda Yung[3]
[1]Eastman School of Music, University of Rochester, Rochester, NY, USA
[2]Brain & Cognitive Sciences, University of Rochester, Rochester, NY, USA
[3]Center for Visual Science, University of Rochester, Rochester, NY, USA

It has been suggested that aesthetic experience is optimal when stimuli have a moderate level of complexity. In this experiment we test this hypothesis, focusing on two musical factors that have not been previously examined in this regard: range and repetition. Our hypothesis is that melodies that are more difficult to encode will seem more complex. Encoding should be inhibited by a large melodic range (since this creates large melodic intervals, which are low in probability) and variability, i.e., lack of repetition (since repetition allows a more parsimonious encoding of the melody).

Melodies were generated by a stochastic process that chooses pitches from a uniform distribution over a defined range. All melodies were 49 notes long and entirely within the G major scale. There were three levels of range: *small* (2 notes), *medium* (6 notes), and *large* (29 notes). There were also three levels of variability: in *low-variability* melodies, each four-note group simply repeated the first (randomly generated) four-note group; in *medium-variability* melodies, each even-numbered four-note group repeated the previous odd-numbered group; and *high-variability* melodies contained no repetition (except that arising by chance). All combinations of range and variability were used. It was predicted that melodies of medium range and variability would obtain the highest ratings, and that there would be an interaction between range and variability such that high-variability melodies were liked more when the range was smaller.

Subjects (musicians and nonmusicians) heard the melodies and indicated their liking for them on a Likert scale. As predicted, in both the range and variability factors, the medium level was preferred over the low and high levels, though the difference between medium and high levels was marginal. The predicted interaction between range and variability was also observed. Implications will be discussed with regard to liking for melodies and melodic encoding.

Musical Persuasion: A Review of Music and Marketing Research

Bradley W. Vines[1]
[1]Nielsen Consumer Neuroscience, London, UK

Background
Music holds a valued place in the marketer's toolset for persuasion, whether in the form of jingles, sonic branding, or background music in stores or in TV ads. Like the application of music in medicine and in education, the application of music in marketing provides a real-world test case for investigating the effects of music on the mind and brain. Though there have been a number of studies investigating music in marketing over the last decades, the most recent review of research on this topic was published back in 1990 (Bruner). We are overdue for a fresh look at this area of research that also takes into account advances in basic research on music.

Aims and Method
This paper develops a contemporary review of research on music and marketing, informed by advances in the study of music from the perspective of psychology and neuroscience. The method involves a thorough literature review of marketing journals as well as scientific journals featuring studies of the potential for music to influence behavior (e.g., purchase decisions).

Results
There are three primary veins of research into music and marketing that have developed. First, studies show that music has a significant impact on consumers in stores. The choice of music relative to the products on sale, the style of music, its tempo, familiarity, and even volume influence how consumers move through the store, as well as what and how much they purchase. Second, the music that accompanies TV advertising impacts not only the interpretation of the storyline in ads, but also the propensity for ads to drive behavior change. Third, by means of its influence on mood, music can alter decision-making and interact with the experience of other senses to color the overall experience of products.

Conclusion
Applied research on the impact of music in marketing confirms the potential of music to influence the human mind and brain. It also opens new avenues of inquiry into multisensory integration and broadens our understanding of the effects of music. This paper sheds light on research investigating music and marketing to date and relates that work to what we currently know from basic research into the psychology and neuroscience of music.

Relationship Between Personality and Music Genres of Macau Youth

Wanfong Viny Hui
Faculty of Education, University of Macau, China
vinyhui@umac.mo

ABSTRACT

The correlation between human personality traits and how people appreciate a specific music genre has been supported by a few Western studies. The present study intended to examine the same issue, with Macau as the setting, by adopting the Big Five Inventory to measure the association between personality and music genres. The secondary focus of this study was to conceptualize how Macau students define the music that they enjoy. A total of 130 Macau undergraduates aged 18–25 participated in the survey, which contained 50 items in the Big Five Inventory and an open-ended item, which asked the participants to name two music genres that they like. Ten music genres (soothing, band, instrumental, folk, jazz, acoustic, R&B, classical, pop, and religious) were identified by manually grouping the responses to the open-ended questions. ANOVA tests were performed to compare genre preference and personality. The participants generally exhibited a wide range of preferences, and their personality traits were more or less mixed, resulting in the finding that no significant differences exist among personality traits and genres. This may be due to the lack of variety in the educational system that they grew up with. According to Spearman's rho correlation analysis of personality traits and genres, no significant correlations exist between the two types of variables. This result echoes the means difference results with the same explanation. As for the open-ended item, as expected, some did not follow the common labels; they used "relax," "soothing," "light" and other terms relating to the effect of music on their moods. These answers may show that some listeners are more concerned about the immediate effect of music on their listening experience and not on the classification of music genres. Therefore, using traditional music labels to reflect the preference of listeners toward a wide and inconsistent system of music genres may be imprecise.

I. INTRODUCTION

Personality is one of the variables in the development of musical preferences that have elicited an increasing amount of research interest. People with different types of personalities react differently toward the same kind of music or musical effects. The stable relationship between musical preference and personality was made evident in a three-year study on adolescents (Delsing, Ter Bogt, Engels, & Meeus, 2008). Early studies (Burt, 1939; Klimas-Kuchtowa, 2000) have reported that extroverts prefer homogenous and sensual music, whereas introverts prefer complex music that sound mystical, introspective, and psychotic (Rawlings et al., 1995).A Polish study (Kopacz, M., 2005) on the relationship between musical elements and Cattell's personality traits revealed several trends. Extroverts who possess such traits as liveliness (carefree and enthusiastic), openness to change (experimenting and freethinking), low vigilance (trusting and forgiving), and social boldness (gregarious and adventuresome) prefer music with melodic themes. In particular, extroverts who exhibit social boldness also prefer music with a faster tempo and meter. Compared with introverts, extroverts prefer a faster tempo. Other musical elements (changes in tempo, dynamics, melodics, leading instrument, and timber) do not have any relationship with personality.

The Big Five personality inventory has been utilized to establish the associations between an individual's favorite music genres and personality. Rentfrow and Gosling (2003) tailored the Short Test of Music Preferences (STOMP) and identified the association of personality traits to musical genres: reflective and complex refer to blues, jazz, classical, and folk music; intense and rebellious refer to rock, alternative, and heavy metal; upbeat and conventional refer to pop, country, sound tracks, and religious music; and energetic and rhythmic refer to rap/hip-hop, soul/funk, and electronic/dance music). Several trends of how different personality traits (openness, extroversion, agreeableness, neuroticism, and conscientiousness) relate to musical preference were established (Rentfrow & Gosling, 2003). Openness was found to be related to reflective and complex music and extroversion to upbeat and conventional music.

Many other revealing trends were established as numerous subsequent studies attempted to replicate and validate the framework of Rentfrow and Gosling (2003). Openness and extroversion exhibit consistent patterns, whereas the other melodic themes and traits have mixed results. Individuals exhibiting openness prefer reflective and complex music, as supported by Brown (2012) in a study on Japanese students; they also prefer intense and rebellious music (Swami, 2013) but dislike upbeat and conventional music. Extroverted individuals prefer all genres, except for reflective and complex. However, Brown's (2012) study on Japanese students and Ercegovac, Dobrota, and Kuscevic's (2015) study on Croatian college students showed that extroverts only like popular music. Individuals exhibiting neuroticism are only interested in upbeat and conventional music and dislike intense and rebellious music. Agreeable individuals prefer upbeat and conventional (Langmeyer et al., 2012) but dislike energetic and rhythmic (Zweigenhaft, 2008). However, agreeableness is a significant predictor for all genres (jazz/world, heavy, and popular) except for the classical genre in Ercegovac, Dobrota, and Kuscevic's study (2015). Only one study has found that people who exhibit conscientiousness (reliability, responsibility, and organization) prefer upbeat and conventional music.

Researchers observed a consistent pattern in the underlying structure of music genres. Delsing et al. (2008) identified rock (heavy metal, punk, Gothic/wave, and rock), elite (jazz, classical, and gospel), urban (hip-hop/rap and soul/R&B), and pop/dance (trance/techno and top 40/charts) in their study on Dutch adolescents. Generally, some of the genres frequently preferred are classical and jazz, rock and heavy metal, and rap/hip-hop. Others also mentioned country music featuring story-telling, new age, and electronic style. Based on previous results on the structure of musical preference, Rentfrow et al.

(2011) further refined five dimensions/factors in the structure of musical preference. They included musical attributes, psychological attributes, and genres to enhance the concept of the multifaceted nature of music styles. The five dimensions are mellow (smooth, relaxing, romantic, R&B, and soft rock), unpretentious (country/singer-songwriter, relaxed, not distorted, and not intelligent), sophisticated (complex, intelligent, inspiring, classical, world, and jazz), intense (loud, forceful, energetic, aggressive, not relaxing, classic rock, and heavy metal), and contemporary (rhythmic, percussive, rap, electronica, funk, and acid jazz).

Various methods of defining music genres for preference measurement have been presented in previous studies. The common method utilized is to simply list a range of genre labels and assess based on a Likert scale. Other studies adopted a highly complicated procedure by playing musical/audio stimuli of a representative genre for a range of selected genres. The first step is to determine the music samples that can fully depict the study subjects' musical preferences. The samples can be audio or verbal descriptors. Several studies looked into the featured musical elements (e.g., tempo, rhythm, melody, and instrument) in musical works to express musical preference (Kopacz, 2005). Preference researchers agree that "musical genres are always contextualized in a particular culture, region and time" (Ferrer, Eerola, & Vuoskoski, 2012, p. 500). To accurately match the genre labels found in commercial or academic arenas with Macau listeners' usage, a certain procedure is implemented to clarify what to put into the preference measurement and how.

The present study serves as a pilot study to investigate the Macau youth's music preference and its relation to personality. The first step is to explore their expression of musical preference: what (genre, label, style, feature, category, etc.) do they use to describe their favorite music and how?

The second step is to identify the differences and correlations between musical preference and the Big Five personality traits.

II. METHOD

A survey was conducted on 130 participants (57 males and 73 females) aged 18 to 25 years old (mean = 20.27, SD = 1.28) who are non-music major undergraduates enrolled in one of the music elective courses in a Macau university. The participants were informed of the purpose of the survey; they expressed their willingness to participate in the study. The survey was administered during the final 10 minutes of each class session.

The questionnaire is composed of a Big Five personality inventory (the 50-item version) translated to Chinese. To obtain information on musical preference, an item asks the respondents to write three of their favorite music genres in free form. The item does not suggest any terms or genres to determine how the participants express their own musical preference.

III. RESULTS

The item on three favorite music genres received 48 terms/labels of music with a total frequency of 242. The researcher consolidated the labels according to psychological expression (soothing, relaxing, etc.), musical characteristic (allegro, light, etc.), genre (heavy metal, folk, etc.), musical period (romantic, modern, etc.), and performing form (choir, orchestral, etc.). Finally, 10 factors of music (Table 1) were obtained. The most popular factor was pop songs (Cantopop, Mandopop, K-pop, J-pop, and English pop), accounting for 64 out of the 242 answers or 26.45%. The second factor was the creation of uplifting music (soothing, relaxing, easy listening, soft, slow, light, new age as used for yoga, allegro, and positive) (19.01%). The third factor was the musical period (classical, romantic, and modern) (14.46%). The fourth factor consisted of band, rock, indie, post hardcore, heavy metal, and alternative rock music (12.40%). The fifth factor included R&B, soul, hip pop, dance, reggae, blues, and rap (10.74%). The sixth factor was folk (folk, oldies, and country music) with 4.55%. The seventh factor covered instrumental music (pure music, orchestral, and piano music) (4.13%). The eighth factor was jazz (jazz and bossa) with 3.31%. The ninth factor covered acoustic music (electronic, anime, and acoustic) (2.89%). The least liked one was choral (hymn, duets, church, chant, and choir) (2.07%).

Table 1. Genre descriptors.

Music	Frequency	Percentage
Pop, K-Pop, J-Pop, Cantopop, English song, Mandopop	64	26.45
Soothing, relax, easy listening, soft, slow, light, New Age, allegro, positive	46	19.01
Classic, Romantic, Modern	35	14.46
Band, rock, indie, post hardcore, heavy metal, alternative rock	30	12.40
R & B, soul, hip-hop, dance, reggae, blues, rap	26	10.74
Folk, oldies, country music	11	4.55
Pure music, instrumental, orchestral, piano	10	4.13
Jazz, bossa	8	3.31
Acoustic, electronic, anime	7	2.89
Hymn, duets, church, chant, choir	5	2.07

The personality score was rated on a seven-point Likert scale (see Table 2). Among the Macau youth participants, agreeableness was the most commonly observed trait (M = 4.66), followed by conscientiousness (M = 4.42), openness (M = 4.24), extroversion (M = 3.96, and emotional stability (M = 3.64).

ANOVA was performed to obtain the differences between the Big Five traits and music preferences. No significant difference was found between all personality traits and music preferences.

Spearman's rho correlation was utilized to analyze any correlation between the five personality traits and music preferences. No significant correlation was found (extroversion: 0.0, agreeableness: −0.09, conscientiousness: −0.10, emotional stability: −0.04, and openness: −0.05).

Table 2. Mean, SD, Max, Min, F Value and P Value of Big Five Personality.

Personality Trait	Mean	SD	Max	Min	F Value	P Value
Agreeableness	4.66	.69	6.13	1.75	.598	.813
Conscientiousness	4.42	.87	7.00	2.00	.622	.793
Openness	4.24	.68	6.70	2.20	1.223	.283
Extraversion	3.96	.78	5.90	1.80	1.460	.163
Emotional stability	3.64	.88	5.78	1.11	.866	.567

IV. DISCUSSION AND CONCLUSION

The raw data on the participants' preferred music indicate that the answers mainly covered the most accessible music, such as pop (in different languages), and functional music to enhance their daily living. The most frequently used descriptors are positive, soothing, and relaxing. Music on the Internet has permeated the lives of young people; many genres, such as heavy metal and jazz, have become user friendly. Young people depend on the Internet to search for and share patterns to create their own listening list. In the old days when musical knowledge was limited to the school, Macau students did not have many choices of music genres to express their preference other than Cantopop, the culture's mainstream genre. Many music labels, such as classical, romantic, and hymns, were probably learned by the participants through the general music education and elective music courses they completed. Future studies can consider modifying the pattern of the five-factor structure of Rentfrow et al. (2011) to design a music preference inventory because it comprises several detailed samples of music using psychological attributes and musical attributes/elements. Instead of asking free-form open-ended questions, future studies should provide musical excerpts for the preference items.

The lack of significant differences and correlations between personality traits and music genres can be explained by the method of preference assessment. Another explanation might be the homogenous education background of the participants. The effectiveness of music training, school culture, and media provision from their 11 years of foundational education can distort the results. Despite widespread Internet use, the training for independent thinking might not have been successful. As exhibited by the distribution of personality traits, agreeable individuals are the majority. According to literature, two traits, namely, extroversion and openness, are robust predictors of music preferences. However, the present study's results do not agree with the trends found in other cultures (Rentfrow and Gosling, 2003).

ACKNOWLEDGMENT

This report, as part of the research project, is supported by the University of Macau.

REFERENCES

Brown, R. A. (2012). Music preferences and personality among Japanese university students. *International Journal of Psychology, 47*(4), 259-268. doi: 10.1080/00207594.2011.631544

Burt, C. (1939). The factorial analysis of emotional traits. *Character and Personality, 7,* 238-290.

Chamorro-Premuzic, T., & Furnham, A. (2007). Personality and music: Can traits explain how people use music in everyday life? *British Journal of Psychology*, *98,* 175-185. doi: 10.1348/000712606X111177

Ercegovac, I. R., Dobrota, S., & Kuščević, D. (2015). Relationship Between Music and Visual Art Preferences and Some Personality Traits. *Empirical Studies of the Arts, 33,* 207-227

Ferrer, R., Eerola, T., & Vuoskoski, J. K. (2013). Enhancing genre-based measures of music preference by user-defined liking and social tags. *Psychology of Music, 41,* 499-518.

Klimas-Kuchtowa, E. (2000). Music preferences and music prevention. *Publication of the Academy of Music in Wroclaw*, *76*

Kopacz, M. (2005). Personality and Music Preferences: The Influence of Personality Traits on Preferences Regarding Musical Elements. *Journal of Music Therapy, 42,* 216-239. doi: 10.1093/jmt/42.3.216

Langmeyer, A., Guglhör-Rudan, A., & Tarnai, C. (2012). What Do Music Preferences Reveal About Personality? *Journal of Individual Differences, 33*(2), 119-130. doi: 10.1027/1614-0001/a000082¨

Litle, P., & Zuckerman, M. (1986). Sensation seeking and music preferences. *Personality and Individual Differences*, *7,* 575-57.

Rawlings, D., Hodge, M., Sherr, D., & Dempsey, A. (1995). Tough-mindedness and preferences for musical excerpts, categories, and triads. *Psychology of Music, 23,* 63-80.

Rentfrow, P. J., Goldberg, L. R., & Levitin, D. J. (2011). The structure of musical preferences: A five-factor model. *Journal of Personality and Social Psychology, 100,* 1139-1157. doi: 1.1037/a0022406

Rentfrow, P. J., & Gosling, S. D. (2003). The do re mi's of everyday life: The structure and personality correlates of music preferences. *Journal of Personality and Social Psychology, 84,* 1236-1256. doi: 10.1037/0022-3514.84.6.1236

Schafer, T., & Sedlmeier, P. (2009). From the functions of music to music preference. *Psychology of Music, 37*, 279-300. doi: 10.1177/0305735608097247

Schwartz, K. D., &Fouts, G. T. (2003). Music preferences, personality style, and developmental issues of adolescents. *Journal of Youth and Adolescence, 32,* 205-213

Swami, V., Malpass, F., Havard, D., Benford, K., Costescu, A., Sofitiki, A., & Taylor, D. (2013). Metalheads: The influence of personality and individual differences on preference for heavy metal. *Psychology of Aesthetics, Creativity, and the Arts, 7*, 377.

Zweigenhaft, R. L. (2008). A Do Re Mi Encore. A closer look at the personality correlates of music preferences. *Journal of Individual Differences, 29,* 45-55. doi: 10.1027/1614-0001.29.1.45

The catchiness phenomenon: genre-specific preferences

Julia I. Klein[1], Carol L. Krumhansl[1]
[1]Department of Psychology, Cornell University, Ithaca, NY, USA

Short, looping segments of very catchy songs that spontaneously come to mind are known as earworms (Williamson, 2012). It has been estimated that over 90% of people experience this type of involuntary musical imagery (INMI) every week (Liikkanen, 2012a). Despite our understanding of the incidence of INMI, less is known about specific underlying musical qualities of earworms, or more broadly, what qualities make a song catchy and likely to become an earworm. This research considered what musical qualities make music catchy, using our knowledge of how familiarity and liking play a role in determining musical preferences. The first study investigated the "catchiness" of the top *Billboard* year-end hits from 1910-2009 and aimed to identify musical features associated with "catchiness". After adjusting for the strong relationship between catchiness and familiarity, three musical factors exhibited the strongest relationship with catchiness: tempo, harmonic complexity (measured as the number of different chords in the chorus/hook), and the incidence of chain rhyming in the chorus/hook. A follow up study was conducted to explore if these three factors still hold in unfamiliar musical stimuli, with a particular focus on genre to determine if these preferences are genre-specific. Results indicated different patterns of preferences across the five genres studied (Country, Hip-Hop, Pop, R&B and Rock). The current study concludes that the musical qualities that make a song catchy, and therefore more likely to become an earworm, tend to make songs easier to sing, making them intrinsically more memorable.

Neurophysiological and Behavioral Measures of Musical Engagement

Blair Kaneshiro,[*1] Duc T. Nguyen,[*] Jacek P. Dmochowski,[#†] Anthony M. Norcia,[#] Jonathan Berger[*2]

[*]*Center for Computer Research in Music and Acoustics, Stanford University, USA*

[#]*Department of Psychology, Stanford University, USA*

[†]*Department of Biomedical Engineering, City College of New York, USA*

[1]blairbo@ccrma.stanford.edu, [2]brg@ccrma.stanford.edu

ABSTRACT

Inter-subject correlations (ISCs) of cortical responses have been shown to index audience engagement with narrative works. In parallel lines of research, continuous self-reports and physiological measures have linked listener engagement and arousal with specific musical events. Here we combine EEG-ISCs, physiological responses, and continuous self-reports to assess listener engagement and arousal in response to a full-length musical excerpt. The temporal resolution of these measures permits connections to be drawn among them, and also to corresponding musical events in the stimulus. Simultaneous 128-channel EEG, ECG, and respiratory inductive plethysmography were recorded from 13 musicians who heard the first movement of Elgar's E-Minor Cello Concerto in original and reversed conditions. Continuous self-reports of engagement with each excerpt were collected in a separate listen. ISCs of EEG responses were computed in the subspace of its maximally correlated component. Temporally resolved measures of activity and synchrony were computed from heart rate, respiratory rate, respiratory amplitude, and continuous behavioral responses. Results indicate that all but one of the response measures (heart rate activity) achieves statistical significance over the course of the original excerpt. Furthermore, regions of statistical significance can be linked to salient musical events. Finally, activity and synchrony of a given physiological or behavioral response at times occur in alternation, highlighting the utility of considering both measures. These findings constitute a first step toward relating EEG-based measures of engagement to physiology and behavior, while also helping to further disentangle engagement and arousal.

I. INTRODUCTION

The enjoyment of music is ubiquitous among humans. Its experience varies in the quality and intensity of engagement with the sound, ranging from passive and subconscious to deeply engrossed and attentive. In addition to audition – whether active or passive – music can serve to organize and coordinate movement, and also serves a variety of other utilitarian and hedonistic functions. As a result, both qualitative and quantitative measures of engagement are elusive. Here we introduce an integrated approach to measure musical engagement through cortical responses, physiological markers associated with arousal and emotional response to music, and continuous self-reports.

In recent years, the use of inter-subject correlations of cortical responses has facilitated the study of neural processing of naturalistic stimuli. In addition to providing a data-driven approach toward identifying brain regions that activate reliably across individuals, this technique has also highlighted the promise of using neural synchrony as a measure of attentive engagement. For example, inter-subject correlations (ISCs) of responses recorded using functional magnetic resonance imaging (fMRI) have been shown to increase during emotionally arousing film scenes (Hasson et al., 2004) and in response to rhetorically powerful (as opposed to rhetorically weak) speeches (Schmälzle et al., 2015). While electroencephalographic (EEG) responses provide superior temporal resolution, on the same time scale as musical events, the low signal-to-noise ratio and high dimensionality of this data modality have historically limited its feasibility in analyzing responses to full-length, naturalistic musical works (Kaneshiro & Dmochowski, 2015). However, a recently developed spatial filtering technique for EEG has been successfully applied in an ISC analysis paradigm to study engagement with naturalistic stimuli. This technique, termed Reliable Components Analysis (RCA), computes maximally correlated components across a collection of EEG records (Dmochowski et al., 2012, 2014; Dmochowski, Greaves, & Norcia, 2015). EEG-ISCs computed in one-dimensional subspaces of maximally reliable components have been shown to reflect suspense, tension, and engagement with audiovisual film excerpts (Dmochowski et al., 2012, 2014). EEG-ISCs have also been shown to be higher for intact excerpts of vocal music than for phase-scrambled controls (Kaneshiro et al., 2014). Importantly, this method permits an experimental paradigm in which EEG responses can be compared across audience members, and stimuli representing full works can be presented to experimental participants in a single-exposure paradigm.

A parallel line of research has explored physiological responses to music under the general assumption that music does, indeed, evoke physiological responses in listeners. Studies have often focused on physiological correlates of emotion or arousal in response to naturalistic music excerpts. Distinct changes in continuous physiological measurements that correspond with particularly noticeable musical events (such as a significant structural demarcation or a violated expectation) are noted. Assessing synchrony of physiological responses, the approach taken with cortical ISCs, has to our knowledge not been attempted in music studies, though a relationship between synchrony of heart rate and skin conductance and subsequent charitable contributions has been demonstrated in the setting of viewing an emotional video excerpt (Bracken et al., 2014).

A variety of physiological responses have been employed to characterize and measure fluctuations in arousal or emotional valence associated with music listening. One response of interest is the galvanic skin response (GSR), which is generally found to increase with arousal (Gomez & Danuser, 2004; Gomez & Danuser, 2007; Rickard 2004), harmonic and melodic unexpectedness (Steinbeis, Koelsch, & Sloboda, 2006; Koelsch et al., 2008; Egermann et al., 2013), and in conjunction with musical 'chills' (Grewe et al., 2010). Respiratory rate has also been found to increase for faster and more arousing musical excerpts (Gomez & Danuser, 2004;

Gomez & Danuser 2007; Russo, Vempala, & Sandstrom, 2013), while heart rate sometimes increased in response to more arousing music (Russo, Vempala, & Sandstrom, 2013), but often produced no statistically significant change from baseline measures (Gomez & Danuser, 2004; Rickard, 2004; Steinbeis, Koelsch, & Sloboda, 2006; Gomez & Danuser 2007; Koelsch et al., 2008).

Ongoing behavioral responses to music along various dimensions, including locus of attention (Madsen & Geringer, 1990), 'aesthetic experiences' (Madsen, Brittin, & Capperella-Sheldon, 1993), and combined valence and arousal measures (Schubert 2004), have been collected using continuous response interfaces. A recent study of particular relevance to the present study involves continuous behavioral ratings of engagement with a live dance performance (Schubert, Vincs, & Stevens, 2013). This study utilizes a definition of engagement that we will use here. The authors additionally consider both the level of engagement and the agreement among participant responses over time. One finding, which we will return to later, is that so-called 'gem moments' – in that study described as moments of surprise – bring about sudden rises in engagement *level*, but periods of heightened *agreement* tend to occur more when audience expectations have been established and are not interrupted.

In the present study, we aim to validate the application of RCA and EEG-ISCs to study cortical responses to a complete and self-contained musical excerpt from the classical repertoire (in this case a movement from a concerto). Our proposed collection of responses will lead to measures intended to index both engagement (EEG, continuous behavioral responses) and arousal (physiological responses). The responses can additionally be categorized as objective (neurophysiological responses) or subjective (continuous behavioral responses). As our stimulus of interest expresses large fluctuations in arousal, we further seek to interpret the cortical responses among both the levels and the synchronicity of physiological and continuous behavioral responses. Our present analysis focuses on specific periods of very high or low arousal within the musical excerpt, as well as with passages of increasing harmonic and melodic tension, and the introductions of principal and salient musical themes.

II. EXPERIMENTAL DESIGN

A. Stimuli

Stimuli for this experiment were drawn from the first movement of Elgar's Cello Concerto in E Minor, Op. 85, composed in 1919. This work has been shown to induce chills in past experiments (Grewe et al., 2007b; Grewe et al., 2010). The version used here is the recording of the 1965 Jacqueline du Pré performance with sir John Barbirolli and the London Symphony Orchestra, considered to be a definitive and influential performance of the piece (Solomon, 2009).

In descriptive music analysis, salient events are often designated as structurally relevant. These events can include climactic 'highpoints' (Agawu, 1984); the introduction and reprise of principal thematic materials; a sudden, generally unexpected pause; and significant changes of timbre or texture, such as a change in orchestration or the entrance of a soloist. Moments of this type in the Barbirolli/Du Pré recording of the Elgar concerto are noted in Table 1.

To prepare the experimental stimuli, we used Audacity software to remix the original digital .m4a audio file of the recording to mono and export to .wav format. We then loaded the .wav file in Matlab and added a linear fade-in and fade-out to the first and last 1000 msec of the piece, respectively. A reversed condition was then created by reversing the stimulus waveform. Finally, we added a second audio channel containing intermittent clicks, which were sent directly to the EEG amplifier for precise time-stamping of the stimuli and were not heard by participants, and saved the outputs as stereo .wav files.

Table 1. Points of interest in Elgar's Cello Concerto. We identify a set of salient musical events in the excerpt, including points at which the primary theme is introduced (A1, A2) and the start of sections of buildup (B1, B2), which later culminate in highpoints of the piece (C1, C2). We additionally identify a point of low arousal (D). Labels correspond to the labels in Figure 3.

Time	Event	Label
1:23	Entrance of cello theme	A1
2:07	Cello theme, building to high point	B1
2:31	Highpoint	C1
3:19	Break in content	D
6:15	Entrance of cello theme	A2
6:39	Orchestral theme, building to high point	B2
7:02	High point	C2

B. Participants

For this experiment we sought healthy, right-handed participants between 18-35 years old, with normal hearing, who were fluent in English, and who had no cognitive or decisional impairments. As formal musical training has been shown to produce enhanced cortical responses to music (Pantev et al., 1998), we sought participants with at least five years of formal training in classical music. Qualifying forms of training included private lessons, AP or college-level music theory courses, and composition lessons. Years of training did not need to be continuous. Because we wished to avoid involuntary motor activations found to occur in response to hearing music played by one's own instrument (Haueisen & Knösche, 2001), we recruited participants who had no training or experience with the cello. Finally, as we are interested in listener engagement, we sought participants who reported that they enjoyed listening to classical music, at least occasionally.

We collected usable data from 13 participants (four males, nine females) ranging from 18-34 years of age (mean 23.08 years). All participants met the eligibility requirements for the experiment. Years of formal musical training ranged from 7-17 years (mean 11.65 years). Eight participants were currently involved in musical activities at the time of their experimental session. Music listening ranged from 3-35 hours per week (mean 13.96 hours). Data from three additional participants were excluded during preprocessing due to gross noise artifacts. This study was approved by Stanford University's Institutional Review Board, and all participants delivered written informed consent prior to their participation in the experiment.

C. Experimental Procedure

The experiment was organized in two blocks. The first block involved the simultaneous EEG and physiological recordings. Here, the participant was instructed to sit still and listen attentively to the stimuli as they played (no other task was performed) while viewing a fixation image on the monitor in front of him. Each stimulus was presented once in random order and was always preceded by a 1-minute baseline recording, during which time pink noise was played at a low volume. At the conclusion of each trial, the participant used a computer keyboard to rate the pleasantness, arousal, interestingness, predictability, and familiarity of the excerpt just heard. We additionally collected a rating of genre exposure after the original stimulus only.

In the second block, continuous behavioral measures of engagement were recorded. We used the definition from the Schubert, Vincs, & Stevens (2013) study: Being 'compelled, drawn in, connected to what is happening, interested in what will happen next.' Participants were shown this definition prior to the trials in this block, and indicated their understanding of the definition and the task before proceeding with the trials. The participant operated a mouse-controlled slider whose position was displayed on the computer monitor. Each stimulus was presented once in random order.

The continuous behavioral block was always second in the experimental session because we did not want the specific definition of engagement presented there to influence participants' listening behavior during the neurophysiological block, which was intended to involve no cognitive task or effort on the part of the participant beyond listening to the stimuli. This of course introduces a potential confound into the present study, but we found it to be the most appropriate design, as we wanted to collect the full set of responses from each participant. We note that Steinbeis, Koelsch, & Sloboda (2006), who also employed separate neurophysiological and continuous behavioral recordings, presented the continuous behavioral block first. However, participants in that study were given a separate task during the latter block (comparing lengths of stimuli), and thus were likely not impacted by the task of the previous block.

D. Stimulus Delivery and Data Acquisition

Experiments for both blocks were programmed using the Matlab Psychophysics Toolbox (Brainard, 1997) with Matlab R2013b on a Dell Inspiron 3521 laptop computer running the Windows 7 Professional operating system. Audio was output from the stimulus computer via USB to a Native Instruments Komplete Audio 6 sound card, from which the audio channel containing the stimulus was sent to a Behringer Xenyx 502 mixer, split, and presented through two magnetically shielded Genelec 1030A speakers. The second audio channel, containing the timing clicks, was sent directly to the EEG amplifier through one pin of a modified DB-9 cable, which also sent stimulus trigger labels from Matlab and key press events from the participant. The slider interface for the continuous behavioral response block was programmed using a custom implementation within the Psychophysics Toolbox scheme.

Neurophysiological responses were recorded using the Electrical Geodesics, Inc. (EGI) GES 300 system. Dense-array EEG was recorded with vertex reference using unshielded HCGSN 110 and 130 nets. ECG responses were recorded using a two-lead configuration with hydrocel Ag-Cl snap electrodes. Respiratory activity was measured using thorax and abdomen belts that plugged into a Z-Rip Belt Transducer Module. The ECG and respiratory sensors, leads, and apparatus were obtained from EGI and were approved for use with their Polygraph Input Box (PIB); the electrode net provided the ground for these inputs. While EGI does not currently provide an apparatus for measuring GSR, we attempted to collect it using a custom apparatus which plugged into the PIB. The electrode net and the PIB connected to the Net Amps 300 amplifier. All responses were recorded simultaneously at a sampling rate of 1 kHz. The following filters were applied to the physiological responses at acquisition:

- ECG and respiratory: Highpass 0.1 Hz, lowpass 100 Hz, notch 60 Hz;
- GSR: Highpass 0.05 Hz, lowpass 3 Hz, notch 60 Hz.

No filtering was applied to the EEG responses at acquisition. Physiological signals and amplified EEG signals were recorded using EGI Net Station software, version 4.5.7, on a Power Mac G5 desktop computer running the OSX operating system, version 10.6.8. Continuous behavioral responses were recorded to the stimulus laptop at a sampling rate of 20 Hz.

III. DATA ANALYSIS

E. EEG Preprocessing and Analysis

Initial preprocessing steps were performed using EGI's Net Station software. EEG recordings were bandpass filtered between 0.3-50 Hz and temporally downsampled by a factor of 8 to a sampling rate of 125 Hz. Data records were then exported to .mat format.

All subsequent analyses were performed in Matlab using a custom software implementation. Individual recordings were annotated, and trials were epoched using the precisely timed click events sent from the secondary audio channel directly to the EEG amplifier. Despite the filtering step described above, we found it necessary to DC-correct and detrend each trial epoch of data. Following these procedures, we retained electrodes 1 through 124 for analysis, excluding electrodes on the face. Bad electrodes were identified and removed from the data frame (i.e., the number of rows in the data decreased). Ocular artifacts were removed from concatenated trials of each EEG recording in a semi-automatic procedure using Extended Infomax ICA (Bell & Sejnowski, 1995; Jung et al., 1998), as implemented in the Matlab EEGLAB Toolbox (Delorme & Makeig, 2004). Following this, any electrode for which at least 10% of voltage magnitudes exceeded 50 μV across the recording was flagged as a bad electrode; if any such electrodes were identified, the preprocessing process was re-started with these channels of data removed. Once no further electrodes were identified in this stage, any electrode for which at least 10% of voltage magnitudes exceeded 50 μV for a given trial were removed from that trial only. As final preprocessing steps, we removed noisy transient spikes from the data in an iterative procedure, setting to NaN any data samples whose magnitude voltage exceeded four standard deviations of its channel's mean power. Next, the removed rows corresponding to bad electrodes were reconstituted with rows of NaNs. Finally, a row of zeros representing the reference channel was added to the data matrix, and the data were converted to average reference.

Data from 3 (out of 16) participants who completed the experiment had to be excluded from further analysis, resulting in the 13 usable datasets. Excluded participants were identified during preprocessing on the basis of gross noise artifacts (20 or more bad electrodes). We also excluded the physiological and behavioral responses collected from these participants so that the comparison of results would always involve the same population of participants.

Cleaned data frames for each stimulus were aggregated across participants into time-by-electrodes-by-participant matrices. RCA was computed over responses to both stimuli using a publicly available Matlab implementation (Dmochowski, Greaves, & Norcia, 2015), according to the procedure described in Dmochowski et al. (2012). This is the RCA output that we use in the computation of ISCs. For visual comparison of the topographies only, we also ran RCA separately for the set of responses corresponding to each stimulus condition.

F. Physiological Response Preprocessing

Over the course of data collection, we discovered that the GSR apparatus was not working correctly. This response is therefore excluded from analysis. It appears that the apparatus also introduced substantial noise into the EEG recordings. However, we used the GSR apparatus with all participants in order to maintain a consistent experimental procedure.

The Net Station waveform tool used to filter and downsample the EEG data would perform only the downsampling step on physiological responses. Therefore, to avoid possible aliasing artifacts that could result from downsampling without a prior lowpass filtering step, we exported the physiological responses at the original sampling rate (1 kHz) and performed all filtering operations in Matlab.

Full recordings of ECG responses (prior to epoching) were lowpass filtered to 50 Hz using a zero-phase 8th-order Chebyshev Type I filter, and then downsampled by a factor of 8. We then epoched the data into baseline and stimulus trials. Each ECG signal was converted to a time-resolved measure of heart rate (HR), in beats per minute, using time intervals between successive R peaks, and then spline interpolated to a sampling rate of 125 Hz. We baseline corrected each HR vector by subtracting out the mean bpm across 25-55 seconds of the corresponding baseline recording.

We found the chest and abdomen respiratory activations to be highly correlated, and focus our analyses here on the chest activations only, which tended to be larger in amplitude. Full recordings of respiratory responses were lowpass filtered using a zero-phase 8th-order Butterworth filter with a cutoff frequency of 1 Hz, and then temporally downsampled by a factor of 8. The data were then epoched into baseline and stimulus trials. While from the ECG responses we computed only HR, we can derive two measures of interest from the respiratory activations: Respiratory *rate* and respiratory *amplitude*. Our first step in computing these measures over time was to employ a peak-finding algorithm to identify the positive and negative peaks in the response. From these, respiratory rate over time (RRate) was computed from the time intervals between successive positive peaks. Respiratory amplitude over time (RAmpl) was computed from successive peak-to-trough and trough-to-peak amplitude differences. Finally, both response vectors were spline-interpolated to a sampling rate of 125 Hz. As with the HR data, these measures were baseline corrected using mean values computed from 25-55 seconds of the respective baseline recording.

Due to possible orienting responses at the start of a trial (Lundqvist et al., 2009), as well as spline-interpolation artifacts at the start and end of the trials, we discarded the first and last 5 seconds of data from all physiological responses.

G. Continuous Behavioral Response Preprocessing

Continuous behavioral (CB) responses were aggregated across participants into a time-by-participants matrix. While the interpolation artifacts mentioned above were not an issue here, this type of response is known to suffer from reliability issues at the start and end of the stimulus (Schubert, 2013). Therefore, to account for that and for any orienting responses, we discarded the first and last 10 seconds of response from further analysis. Once these potentially transient portions of the data were removed, we z-scored each response.

H. Inter-Subject Correlations

Inter-subject correlations of EEG, HR, RAmpl, RResp, and CB responses were each computed using a 10-second sliding window that advanced in 1-second increments, resulting in an effective temporal resolution of 1 Hz. In each time window, ISCs were computed across all unique pairs of participants; for 13 participants, this comes out to n-choose-13=78 pairs. We present mean ISCs across participant pairs for every temporal window, mapped to the midpoint times of the windows.

I. Statistical Analyses

We performed paired two-tailed t-tests on the participant ratings that were delivered after each trial of the neurophysiological block, using the Bonferroni correction for multiple comparisons (McDonald, 2014).

To assess the statistical significance of mean HR, RAmpl, RRate, and CB activity responses, we performed two-tailed Wilcoxon signed-rank tests (Lehmann, 2006) across the collection of responses at every time point. This assessed whether the sample of responses at a given time point was drawn from a zero-median distribution. We report statistically significant results at $\alpha=0.05$ after controlling for False Discovery Rate (Benjamini & Yekutieli, 2001).

Statistical significance of time-resolved ISCs was assessed using a permutation test (Fisher, 1971). In every permutation iteration, each participant's response vector was shuffled according to non-overlapping 5-second segments. ISCs were then computed over the collection of the shuffled responses. This procedure was repeated for 500 iterations. The 0.95 quantile of mean ISCs across all permutation iterations designates the statistical significance threshold.

IV. RESULTS

J. Behavioral Ratings

We first analyzed the behavioral ratings collected during the neurophysiological block to determine whether participant ratings of pleasantness, arousal, interestingness, predictability, and familiarity differed significantly according to stimulus condition. The distribution of responses, along with responses from individual participants and p-values from the tests, are shown in Figure 1. Using the Bonferroni-corrected p-value threshold of 0.01, only the dimensions of pleasantness and predictability are found to vary significantly according to the stimulus condition. In both cases, the original (forward) stimulus garners higher ratings. Interestingly, we note that while familiarity with the stimuli was low overall, two of the 13 participants reported that they were at least moderately familiar with the reversed stimulus. Finally, ratings of genre exposure range from 2-9 with a mean value of 5.08, verifying that all participants met the inclusion criterion of listening to classical music at least occasionally.

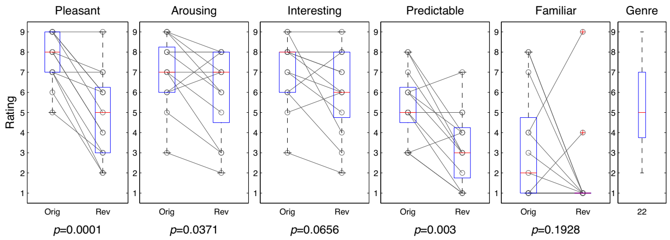


Figure 1. Behavioral ratings of the stimuli. Participants delivered ratings along dimensions of pleasantness, arousal, interestingness, predictability, and familiarity after hearing each stimulus for the first time. Participants reported their exposure to the genre of the original stimulus only. Boxplots as well as individual participant ratings are shown. Ratings of Pleasantness and Predictability are found to differ significantly by stimulus condition using a corrected *p*-value threshold of 0.01.

K. RC Topographies

The RCA algorithm returns a mixing matrix W, which is used to project the EEG data from electrode space to component space. The scalp projections of the weights, useful for visualizing the topographies of components, can be computed by means of a forward model (Parra et al., 2005; Dmochowski et al., 2012). The projections corresponding to the weights of the most reliable component (RC1) are shown in Figure 2. As shown in the figure, a similar fronto-central topography emerges whether RCA is computed over the responses to both stimuli together, or over the responses to the original or reversed stimulus only. While the present topographies are roughly consistent with the RC1 topography produced in response to intact musical excerpts by Kaneshiro et al. (2014), they are now somewhat less symmetric and smooth, which we believe was due to the added noise introduced to the EEG recordings by the GSR apparatus.

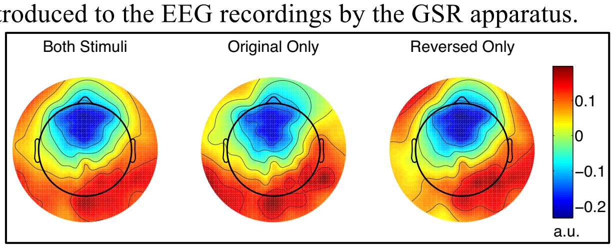


Figure 2. Scalp topographies of maximally reliable EEG components. Left: RC1 was computed over responses to both stimuli. Center, Right: RC1 was computed separately over responses to each stimulus condition. Topographies are roughly consistent with past RC1 topographies in response to music.

L. Combined Responses

Our present focus for the remaining results concerns responses to the original stimulus only. To summarize the neurophysiological and continuous behavioral measures that were derived from the responses recorded across the two experimental blocks, we may assess the following measures in tandem: EEG-ISCs, CB activity, CB-ISCs, HR activity, HR-ISCs, RAmpl activity, RAmpl-ISCs, RRate activity, and RRate ISCs. While we do not attempt to relate the responses to one another quantitatively at this time, several interesting preliminary insights emerge from visual inspection of the responses, in particular regarding their respective regions of statistical significance, in relation to points of interest in the stimulus that we defined in Table 1.

The responses are shown in aggregate in Figure 3. The gray line plots in the top portion of the figure show scale-free activations or ISCs of the responses over the course of the musical excerpt. The colored regions denote points at which the response was deemed statistically significant. The bottom portion of the figure shows the waveform of the stimulus and the stimulus spectrogram. The vertical dotted lines denote the time points of interest as described in Table 1.

As can be seen from the plot, all responses except HR activity produce statistically significant activity or synchrony at some point over the course of the excerpt. The EEG-ISCs, our primary measure of interest, produce the greatest proportion of statistically significant results. Following that, CB activity and CB-ISCs achieve the greatest proportions of statistical significance across the excerpt.

In relation to our identified points of interest throughout the piece, it appears that different measures are significant at different times. The first entrance of the cello theme (A1) brings about significant CB activity, as well as RRate-ISCs, which, from inspection of the RRate plot above it, appear linked to an overall decrease in respiratory rate. The second entrance of the principal theme (A2) is accompanied by significant EEG-ISCs; while CB activity is not significant here, HR-ISCs are significant for a short time. The period of rising tension beginning at B1 and culminating in the first highpoint at C1 implicates EEG-ISCs and CB activity, with the onset of the highpoint itself co-occurring with significant CB activity, CB-ISCs, RAmpl-ISCs, and RRate-ISCs. Somewhat similar observations can be made in the buildup to the second highpoint, between B2 and C2. Interestingly, point D, which we identified as a point of interest due to its *absence* of activity, corresponds with a significant region of RAmpl activity, and is followed closely by significant CB-ISCs, HR-ISCs, and RAmpl-ISCs.

It appears that for the responses from which we derived dual activity and ISC measures (all responses except EEG), significant regions of activity do not necessarily coincide with significant regions of synchrony. In particular, the continuous behavioral responses appear to alternate in significance, and in fact point back to the finding reported by Schubert, Vincs, & Stevens (2013) regarding ‘gem moments’. Based on the activations around C1 and C2, it appears that for the present experiment, the ‘gem moments’ might be the highpoints of the piece, at which time the continuous behavioral activity is significantly above baseline. It may also be the case that significant CB-ISCs relate to the building of tension around many of the structural segmentation points of interest.

V. DISCUSSION

In this study, we have further validated the use of a state-of-the-art analysis method to study cortical responses to naturalistic music excerpts presented in a single-listen

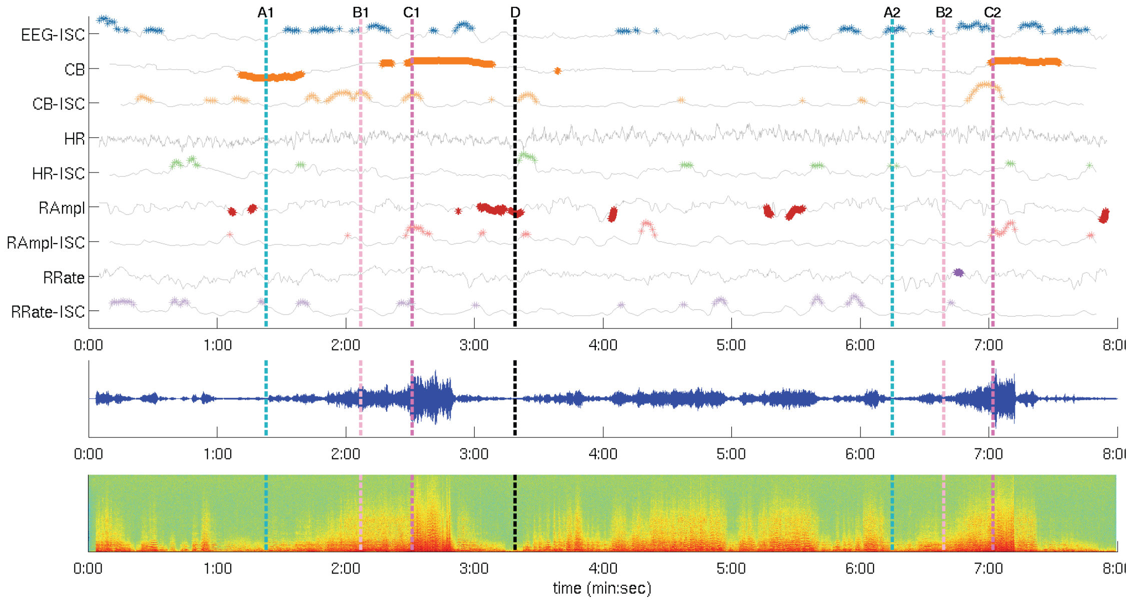


Figure 3. Aggregate neurophysiological and continuous behavioral measures. Top: The measures, in order, are EEG inter-subject correlations (EEG-ISCs), continuous behavioral median z-scored activity and inter-subject correlations (CB, CB-ISC), heart rate median deviation from baseline and inter-subject correlations (HR, HR-ISC), respiratory amplitude median deviation from baseline and inter-subject correlations (RAmpl, RAmpl-ISC), and respiratory rate median deviation from baseline and inter-subject correlations (RRate, RRate-ISC). The colored regions on each line plot denote periods of statistically significant activity or synchrony. Middle/Bottom: The stimulus waveform and spectrogram are plotted for reference. Dashed vertical lines correspond to musical events of interest. Annotations at the top of the plot (A1, B1, C1, D, A2, B2, C2) correspond to the labels in Table 1.

paradigm using EEG. Furthermore, we took first steps toward analyzing these responses alongside various physiological and behavioral responses, seeking points of correspondence among the responses and in relation to the musical stimulus. We analyzed both the activity and the synchrony of the physiological and continuous behavioral responses, and found that these dual measures reached statistical significance at times in tandem, at times in alternation. Finally, by assessing measures intended to denote engagement in conjunction with measures intended to denote arousal, we took a first step toward disentangling these two constructs.

A number of issues can be improved upon in future iterations of this research. First, more work is needed to integrate a GSR recording modality into the EGI PIB for simultaneous collection with the EEG responses. Next, given that our stimulus was chosen for its variation in arousal, which is arguably related to loudness, it may be useful to introduce other control stimuli that retain the amplitude envelope of the original version while manipulating the musicality and temporal coherence of the underlying content. Finally, it is important to note that each response analyzed here occurs at some point after its corresponding musical event; these lags vary according to the measure and have not been definitively determined. For example, while we may assume that EEG responses occur within 500 msec of a stimulus and, in the context of the 10-second ISC window, can be treated as concurrent with the stimulus, more care must be taken with the other responses. Stimulus-to-response time has been estimated to range from 1-5 seconds for physiological responses (Schubert & Dunsmuir, 1999; Grewe et al., 2007a; Grewe, Kopiez, & Altenmüller, 2009; Bracken et al., 2014), and roughly 1-3 seconds for behavioral responses (Krumhansl, 1996; Schubert, 2004; Sammler et al., 2007; Egermann et al., 2013). From inspection of the regions of significance in Figure 3, it appears that shifting the continuous behavioral and physiological responses back in time by such amounts (to the corresponding stimulus events), might in fact improve their alignment with our musical events of interest in some cases, and lead to alternate interpretations in others. Thus, this will be an interesting topic to consider further in future research.

In conclusion, EEG-ISCs, physiological responses, and continuous self-reports suggest a promising combined framework with which to characterize musical engagement. Fluctuations in the quantitative physiological measures reflect fluctuation in arousal levels, which, when changing in approximate simultaneity across measures, and correspondent with changes in the musical signal (such as changes in spectral energy or intensity, or a rise in musical tension), can be interpreted in conjunction with both cortical and continuous behavioral measures of engagement.

ACKNOWLEDGMENT

This research was supported by a grant from the Wallenberg Network Initiative: Culture, Brain Learning. The authors thank Jay Kadis for his assistance with the custom GSR apparatus.

REFERENCES

Agawu, V. K. (1984) Structural ‘Highpoints’ in Schumann’s ‘Dichterliebe’. *Music Analysis*, *3*(2) 159-180.

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, *29*(4), 1165-1188.
Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural computation*, *7*(6), 1129-1159.
Bracken, B. K., Alexander, V., Zak, P. J., Romero, V., & Barraza, J. A. (2014). Physiological synchronization is associated with narrative emotionality and subsequent behavioral response. *In Foundations of Augmented Cognition,* 3-13. Springer.
Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433-436.
Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, *134*(1), 9-21.
Dmochowski, J. P., Sajda, P., Dias, J., & Parra, L. C. (2012). Correlated components of ongoing EEG point to emotionally laden attention–a possible marker of engagement? *Frontiers in human neuroscience*, *6*(112).
Dmochowski, J. P., Bezdek, M. A., Abelson, B. P., Johnson, J. S., Schumacher, E. H., & Parra, L. C. (2014). Audience preferences are predicted by temporal reliability of neural processing. *Nature communications*, *5*.
Dmochowski, J. P., Greaves, A. S., & Norcia, A. M. (2015). Maximally reliable spatial filtering of steady state visual evoked potentials. *Neuroimage*, *109*, 63-72.
Egermann, H., Pearce, M. T., Wiggins, G. A., & McAdams, S. (2013). Probabilistic models of expectation violation predict psychophysiological emotional responses to live concert music. *Cognitive, Affective, & Behavioral Neuroscience 13*, 533-553.
Fisher, R. A. (1971). *The Design of Experiments*. New York: Macmillan Publishing Co.
Gomez, P., & Danuser, B. (2004). Affective and physiological responses to environmental noises and music. *International Journal of Psychophysiology 53*, 91-103.
Gomez, P., & Danuser, B. (2007). Relationships between musical structure and psychophysiological measures of emotion. *Emotion 7*(2), 377-387.
Grewe, O., Nagel, F., Kopiez, R., & Altenmüller, E. (2007a). Emotions over time: Synchronicity and development of subjective, physiological, and facial affective reactions to music. *Emotion 7*(4), 774-778.
Grewe, O., Nagel, F., Kopiez, R., & Altenmüller, E. (2007b). Listening to music as a re-creative process: Physiological, psychological, and psychoacoustical correlates of chills and strong emotions. *Music Perception*, *24*(3), 297-314.
Grewe, O., Kopiez, R., & Altenmüller, E. (2009). The chill parameter: Goose bumps and shivers as promising measures in emotion research. *Music Perception*, *27*(1), 61-74.
Grewe, O., Katzur, B., Kopiez, R., & Altenmüller, E. (2010). Chills in different sensory domains: Frisson elicited by acoustical, visual, tactile and gustatory stimuli. *Psychology of Music 39*(2), 220-239.
Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science*, *303*(5664), 1634-1640.
Haueisen, J., & Knösche, T. R. (2001). Involuntary motor activity in pianists evoked by music perception. *Journal of cognitive neuroscience*, *13*(6), 786-792.
Jung, T. P., Humphries, C., Lee, T. W., Makeig, S., McKeown, M. J., Iragui, V., & Sejnowski, T. J. (1998). Extended ICA removes artifacts from electroencephalographic recordings. *Advances in neural information processing systems*, 894-900.
Kaneshiro, B., Dmochowski, J. P., Norcia, A. M., & Berger, J. (2014). Toward an objective measure of listener engagement with natural music using inter-subject EEG correlation. In *ICMPC13-APSCOM5*.
Kaneshiro, B., & Dmochowski, J. P. (2015). Neuroimaging methods for music information retrieval: Current findings and future prospects. In *ISMIR'15.*
Koelsch, S., Kilches, S., Steinbeis, N., & Schelinski, S. (2008). Effects of unexpected chords and of performer's expression on brain responses and electrodermal activity. *PLoS ONE* 3:7, e2631.
Krumhansl, C. L. (1996). A perceptual analysis of Mozart's Piano Sonata K. 282: Segmentation, tension, and musical ideas. *Music Perception*, *13*(3), 401-432.
Lehmann, E. L. (2006). *Nonparametrics: Statistical Methods Based on Ranks*. Springer.
Lundqvist, L. O., Carlsson, F., Hilmersson, P., & Juslin, P. (2009). Emotional responses to music: Experience, expression, and physiology. *Psychology of Music 37*(1), 61-90.
Madsen, C. K., & Geringer, J. M. (1990). Differential patterns of music listening: Focus of attention of musicians versus nonmusicians. *Bulletin of the Council for Research in Music Education*, 45-57.
Madsen, C. K., Brittin, R. V., & Capperella-Sheldon, D. A. (1993). An empirical method for measuring the aesthetic experience to music. *Journal of Research in Music Education*, *41*(1), 57-69.
McDonald, J. H. (2014). *Handbook of Biological Statistics*. Baltimore: Sparky House Publishing.
Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., & Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature*, *392*(6678), 811-814.
Parra, L. C., Spence, C. D., Gerson, A. D., & Sajda, P (2005). Recipes for the linear analysis of EEG. *NeuroImage, 28*(2), 326-341.
Rickard, N. S. (2004). Intense emotional responses to music: A test of the physiological arousal hypothesis. *Psychology of Music 32*(4), 371-388.
Russo, F. A., Vempala, N. N., & Sandstrom, G. M. (2013). Predicting musically induced emotions from physiological inputs: linear and neural network models. *Frontiers in Psychology, 4*, article 468.
Sammler, D., Grigutsch, M., Fritz, T., & Koelsch, S. (2007). Music and emotion: electrophysiological correlates of the processing of pleasant and unpleasant music. *Psychophysiology*, *44*(2), 293-304.
Schmälzle, R., Häcker, F. E., Honey, C. J., & Hasson, U. (2015). Engaged listeners: shared neural processing of powerful political speeches. *Social cognitive and affective neuroscience*, nsu168.
Schubert, E., & Dunsmuir, W. (1999). Regression modelling continuous data in music psychology. *Music, mind, and science*, 298-352.
Schubert, E. (2004). Modeling perceived emotion with continuous musical features. *Music Perception: An Interdisciplinary Journal*, *21*(4), 561-585.
Schubert, E. (2013). Reliability issues regarding the beginning, middle and end of continuous emotion ratings to music. *Psychology of music*, *41*(3), 350-371.
Schubert, E., Vincs, K., & Stevens, C. J. (2013). Identifying regions of good agreement among responders in engagement with a piece of live dance. *Empirical Studies of the Arts*, *31*(1), 1-20.
Solomon, J. (2009). Deconstructing the definitive recording: Elgar's Cello Concerto and the influence of Jacqueline du Pré. Unpublished manuscript.
Steinbeis, N., Koelsch, S., & Sloboda, J. A. (2006). The role of harmonic expectancy violations in musical emotions: Evidence from subjective, physiological, and neural responses. *Journal of Cognitive Neuroscience 18*(8), 1380-1393.

Freestyle lyricism expertise in auditory rhyme processing

Keith Cross[1], Takako Fujioka[2]
[1]Graduate School of Education, Stanford University, Stanford, CA, USA
[2]Department of Music, Stanford University, Stanford, CA, USA

This study uses electroencephalography (EEG) to determine how neural functions for auditory rhyme processing are related to expertise in *freestyle lyricism*, a form of Hip-Hop rap improvisation involving the spontaneous composition of rhyming lyrics. Previous event-related potential studies have shown that rhyming congruency is reflected in the phonological N400 component, and that musical expertise affected a difference between technical and aesthetic judgments for musical stimuli, as reflected in the contingent negative variation (CNV) component before the critical moment. Here, we asked expert freestyle lyricists and laypersons to listen to pseudo-word triplets in which the first two words always rhymed while the third one either fully rhymed (e.g., ST**EEK**, PR**EEK**; FL**EEK**), half-rhymed (e.g., ST**EEK**, PR**EEK;** FR**EE***T*), or did not rhyme at all (e.g., ST**EEK**, PR**EEK;** Y*AME*). For each trial, upon hearing the third word, participants had to answer one of two types of questions: one aesthetic (do you 'like' the rhyme?) and one descriptive (is the rhyme 'perfect'?). Responses were indicated by a button-press (yes/no). Rhyme effect was investigated as the difference between full-rhyme and non-rhyme conditions, and was measured beginning from the vowel onset of the critical word. Rhyme effects were observed as a fronto-central negativity, strongest in left and right lateral sites in lyricists; this effect was highly left-lateralized in laypersons. A stronger CNV to the 'perfect' task towards the third word was observed for both lyricists and laypersons, which suggests that determining whether rhymes are perfect requires more exertion than making aesthetic judgments of rhyme. This task effect was observed in lyricists over fronto-central electrode sites while appearing at the midline for laypersons. The results indicate that expertise in freestyle lyricism may be associated with more expansive and bilateral neurocognitive networks for processing of rhyme information.

Investigating the Timecourse of Aesthetic Judgments of Music

Amy M. Belfi,[*1] Jess Rowland,[*2] Edward A. Vessel,[#3] G. Gabrielle Starr[&4], & David Poeppel[*#5]
[*]*Department of Psychology, New York University, USA*
[#]*Department of Neuroscience, Max Planck Institute for Empirical Aesthetics, Germany*
[&]*Department of English, New York University, USA*
[1]amy.belfi@nyu.edu, [2]jr2796@nyu.edu, [3]ed.vessel@aesthetics.mpg.de, [3]gabrielle.starr@nyu.edu, [3]david.poeppel@nyu.edu

ABSTRACT

Here we sought to investigate the timescale of aesthetic judgments of music. In Experiment 1, we used a gating paradigm to measure the amount of time needed for listeners to make reliable aesthetic judgments. Participants heard musical excerpts of increasing duration (250 ms, 500 ms, 750 ms, 1000 ms, 2000 ms, and entire excerpt [10s]). Excerpts consisted of music from three genres: classical, jazz, and electronic. After each excerpt, participants rated how much they liked the excerpt (1-9 point scale). Participants made reliable liking judgments as early as 500 ms and, overall, showed a stronger preference for classical music over electronic. In Experiment 2, we controlled for the amount of musical information conveyed, so that excerpts with faster tempos would not include more musical information than excerpts with slower tempos. Participants heard musical excerpts of increasing duration (the length of one eighth note, quarter note, half note, dotted-half note, whole note, and entire excerpt). The results show that, when controlling for the amount of information, listeners were able to make reliable liking judgments at the shortest stimulus duration (i.e., eighth note). However, this result was driven by the electronic music. We suggest that aesthetic judgments of liking can be made very quickly and, plausibly, that these judgments may differ based on the genre of music.

I. INTRODUCTION

Previous research has shown that listeners can accurately judge familiarity and emotional quality of musical excerpts on a very short timescale (i.e., hundreds of milliseconds; Filipic, Tillmann, & Bigand, 2010; Tillmann, Albouy, Caclin, & Bigand, 2014). Similarly, neuroimaging research has suggested that aesthetic judgments of visual artwork may occur within the first 1000 ms following stimulus onset (Cela-Conde et al., 2013). In the present study, we sought to investigate the timescale of aesthetic judgments of musical excerpts. We also investigated whether there were differences between musical genres in the timecourse of aesthetic judgments.

II. EXPERIMENT 1

A. Method

1) Participants. Twenty students from New York University participated in this study. The data from four participants were excluded (see Data Quantification and Analysis), leaving a total of sixteen participants. Participants had, on average, six years of formal musical training (range: 1–18 years). Participants were undergraduate students enrolled in Psychology courses at NYU and received credit for participating in the current study. All procedures were approved by the New York University Institutional Review Board and participants provided written informed consent before participating.

2) Materials. Musical excerpts consisted of three genres, referred to here as "classical," "jazz," and "electronic." Twenty 10-second excerpts were selected in each genre, totaling 60 excerpts. Within each genre, pieces were selected to be fairly homogenous and stylistically consistent in order to prevent participants from responding purely by genre or subgenre considerations. For classical music, all excerpts were chosen from 19th century small ensemble music of the Romantic era. Nineteeth-century classical was selected because it tends to have a wider range of dynamic and emotional intensity than earlier Classical-era music, and would therefore allow for a wider range of behavioral responses. No composer was included more than once in the stimulus set, and no highly well-known composers were selected. All pieces were from the European tradition. Jazz music consisted of 1960's music with bop or post-bop elements, avoiding avant-garde styles of the later 60s. Ensembles and instrumentation was all traditional jazz, consisting of piano, guitar, saxophone, brass, etc. Electronic music consisted of electronic dance music with a distinctive beat structure (i.e., no ambient music) that contained no lyrics or human vocalizations. All electronic pieces were contemporary and ranged in tempo from 60-150 beats per minute.

3) Procedure. Participants heard the 60 musical excerpts blocked by genre. Within each genre block, each of the 20 excerpts was presented in six sub-blocks of increasing duration (250ms, 500ms, 750ms, 1000ms, 2000ms, 10s). After each excerpt, participants rated how much they liked the excerpt on a 1-9 point Likert scale. The excerpts were presented in a different random order for each participant and genre block order was counterbalanced across participants.

B. Data Quantification and Analysis

For each participant, ratings on the full 10 sec excerpts were used to classify each excerpt as overall 'liked' or 'disliked.' If the participant's rating of the full excerpt was above the median rating of all the full excerpts, it was classified as 'liked,' otherwise it was classified as 'disliked.' Four participants were excluded from analysis because they failed to provide any deviation from their median responses for at least one of the three genres (i.e., these participants reported the same rating for every excerpt in at least one genre).

Liking ratings were analyzed using a 2x3x5 ANOVA with preference (liked, disliked), genre (classical, jazz, electronic) and duration (250, 500, 750, 1,000, and 2,000 ms) as within-subjects factors.

C. Results

The ANOVA revealed a significant main effect of preference on liking ratings [$F(1,15) = 29.82$, $p < 0.001$, $\eta^2 = 0.66$]; excerpts classified as 'liked' were rated higher (M=5.10, SEM = 0.24) than those classified as 'disliked' (M=4.36, SEM = 0.24). There was also a significant main effect of genre [$F(2,30) = 3.87$, $p = .03$, $\eta^2 = 0.20$]; classical (M=5.03, SEM = 0.28) excerpts were rated higher than electronic excerpts (M=4.34, SEM = 0.30), but there were no differences between jazz excerpts (M=4.82, SEM = 0.22) and the othergenres. Lastly, there was a main effect of duration [$F(4,60) = 22.25$, $p < 0.001$, $\eta^2 = 0.59$; longer durations were rated higher than shorter durations.

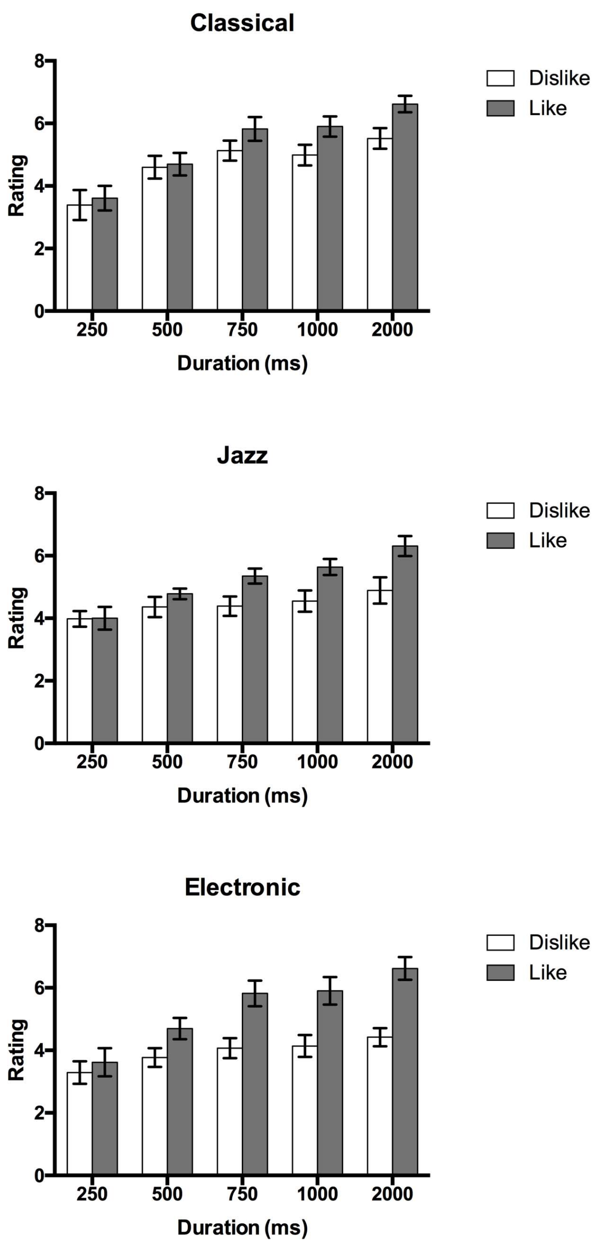


Figure 1. Average liking ratings (minimum= 1, maximum = 9) for excerpts classified as 'liked' or 'disliked' at each duration in Experiment 1. Error bars denote standard error of the mean.

There was a significant interaction between preference and duration [$F(4,69) = 13.55$, $p < 0.001$, $\eta^2 = 0.47$]. Pairwise comparisons revealed that liked excerpts were rated higher than disliked excerpts at the 500 ms duration ($p = 0.01$) and all longer durations (*ps* < 0.01), but not at the 250 ms duration ($p = 0.20$).

D. Discussion

Listeners can make reliable aesthetic judgments as early as 500 ms after the onset of a musical stimulus. This timescale is similar for other judgments, such as familiarity and emotional content. These results also suggest a higher preference overall for classical and jazz music compared to electronic.

III. EXPERIMENT 2

Experiment 1 revealed the amount of time listeners' need to make an accurate aesthetic judgment of music (~500ms). It is unclear from these observations how much *information* is needed to make this type of judgment. An excerpt with a slower tempo may convey less musical information in 250 ms than an excerpt with a faster tempo. To control for this potential difference, we conducted Experiment 2. The procedure was identical to Experiment 1, except duration blocks were not set by absolute time but instead by note values based on the tempo of each piece.

A. Method

1) Participants. Twenty students from New York University participated in this study. Participants in Experiment 2 were different from those in Experiment 1. Four participants were excluded, leaving a total of sixteen participants. Participants had, on average, six years of formal musical training (range: 0–12 years). Participants were undergraduate students enrolled in Psychology courses at NYU and received credit for participating in the current study. All procedures were approved by the New York University Institutional Review Board, and participants provided written informed consent before participating.

2) Materials. Materials are as in Experiment 1. However, instead of designating duration blocks of absolute time, as in Experiment 1, excerpts in Experiment 2 were played for durations designated by musical information. That is, excerpts were blocked by the amount of information (i.e., notes) as opposed to absolute time. Blocks were as follows: eighth note, quarter note, half note, dotted half note, whole note, and entire excerpt. On average, the absolute durations for these were as follows: 260 ms, 520 ms, 1000 ms, 1,500 ms, 2,000 ms. A one-way MANOVA revealed significant differences between genres in terms of duration [$F(2,57) = 4.61$, $p = 0.01$, $\eta^2 = 0.13$]. All durations of electronic music were significantly longer than durations of jazz music (*ps* < 0.01). See **Table 1** for mean duration values for all three genres.

Table 1. Mean duration values for each genre in Experiment 2.

Durations	Genres		
	Classical	Jazz	Electronic
8th note	270ms	220ms	280ms
Quarter note	540ms	440ms	560ms
Half note	1080ms	900ms	1120ms
¾ note	1620ms	1340ms	1680ms
Whole note	2160ms	1800ms	2240ms

3) Procedure. Participants heard the 60 musical excerpts blocked by genre. Within each genre block, each of the 20 excerpts was presented in six sub-blocks of increasing duration (see above). After each excerpt, participants rated how much they liked the excerpt on a 1-9 point Likert scale. excerpts were presented in a different random order for each participant and genre block order was counterbalanced across participants. Data were analyzed in the same manner described in Experiment 1.

B. Results

The ANCOVA revealed a significant main effect of preference [$F(1,15) = 118.30$, $p < 0.001$, $\eta^2 = 0.88$]; excerpts classified as 'liked' were rated higher (M=5.62, SEM = 0.30) than those classified as 'disliked' (M=4.20, SEM = 0.32). There was also a main effect of duration [$F(4,60) = 12.50$, $p < 0.001$, $\eta^2 = 0.45$; longer durations were rated higher than shorter durations. There was no main effect of genre [$F(2,30) = 1.04$, $p = .01$, $\eta^2 = 0.06$].

There was a significant interaction between preference and duration [$F(4,60) = 24.26$, $p < 0.001$, $\eta^2 = 0.61$]. Pairwise comparisons revealed that 'liked' excerpts were rated higher than 'disliked' excerpts at all durations, including the shortest duration (*ps* < 0.01). There was a significant interaction between preference and genre [F(2,30)=18,40, p < 0.001, $\eta^2 = .50$]. Pairwise comparisons indicated that the 'disliked' excerpts were rated significantly lower for electronic (M = 3.65, SEM = 0.32) than the jazz (M = 4.37, SEM = 0.36) and classical excerpts (M = 4.58, SEM = 0.37). There were no genre differences for ratings of the 'liked' excerpts. There was a trending three-way interaction between preference, genre, and duration [$F(8,120) = 1.83$, p = 0.07, $\eta^2 = 0.11$]. Liked and disliked excerpts were rated significantly differently at the shortest duration for electronica, but were not distinguished until the third duration for classical and jazz (**Figure 2**).

C. Discussion

Compared to Experiment 1, these results indicate that listeners make reliable aesthetic judgments at the shortest duration, which on average was 220 ms. When we controlled for the amount of information present at each duration, there was no longer a main effect of genre. That is, the preference for classical over electronica seen in Experiment 1 seems to be due to differences in the amount of information conveyed.

IV. CONCLUSION

Listeners can make aesthetic judgments of music relatively quickly (i.e., within hundreds of milleseconds). Here we show that listeners can make aesthetic judgments even more quickly when the amount of musical information is controlled. Interestingly, a stronger preference for 'liked' excerpts as duration increases seems to be driving this effect. That is, ratings for 'liked' excerpts increase as duration increases, as opposed to ratings for 'disliked' excerpts decreasing. While, for all genres tested here, aesthetic judgments of liking can be made based on very sparse information, our results suggest that that these judgments may differ based on the genre of music. For certain genres, judgments of aesthetic preference may be made on a similar timescale to judgments of familiarity or emotional content.

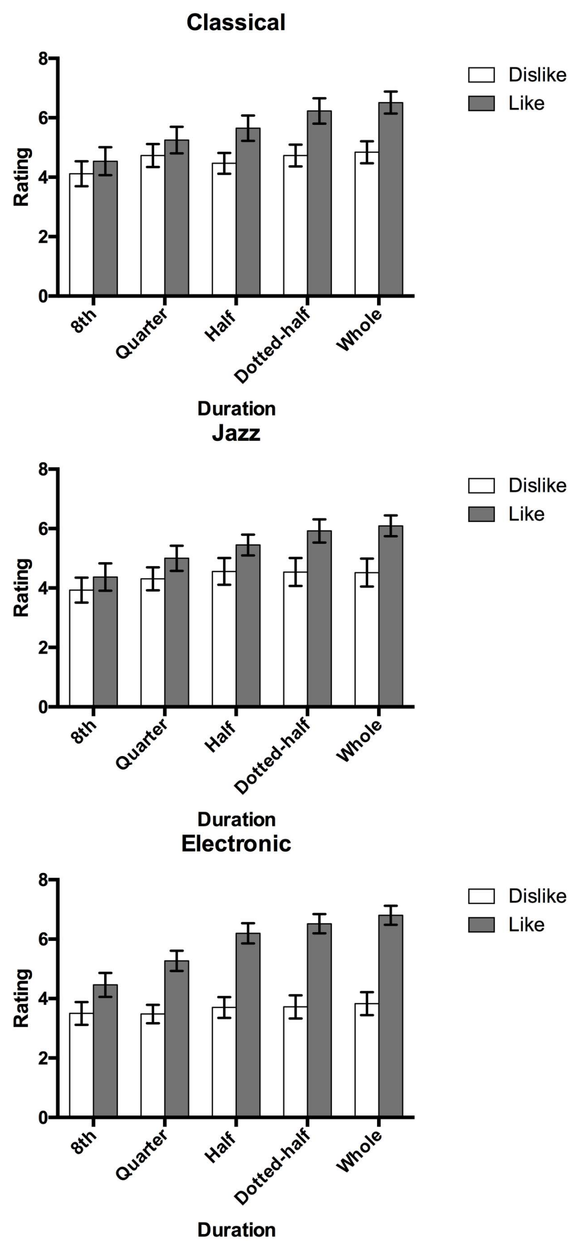


Figure 2. Average liking ratings (minimum= 1, maximum = 9) for excerpts classified as 'liked' or 'disliked' at each duration of Experiment 2. Error bars depict standard error of the mean.

ACKNOWLEDGMENT

We would like to thank research assistants Yvonne Bowman and Kayla Lim for their help collecting data.

REFERENCES

Cela-Conde, C. J., Prieto-Garcia, J., Ramasco, J. J., Mirasso, C. R., Bajo, R., Munar, E., … Maestu, F. (2013). Dynamics of brain networks in the aesthetic appreciation. *Proceedings of the National Academy of Sciences, 110,* 10454-10461.

Filipic, S., Tillmann, B., & Bigand, E. (2010). Judging familiarity and emotion from very brief musical excerpts. *Psychonomic Bulletin & Review, 17,* 335-341.

Tillmann, B., Albouy, P., Caclin, A., & Bigand, E. (2014). Musical familiarity in congenital amusia Evidence from a gating paradigm. *Cortex, 59,* 84-94.

Gesture-sound causality from the audience's perspective: investigating the influence of mapping perceptibility on the reception of new digital musical instruments

Gina Emerson[1], Alexander Förstel[2] and Hauke Egermann[3]
[1]Musicology Insitute, Humboldt Universität, Berlin, Germany
[2]Audio Communication Group, Technische Universität, Berlin, Germany
[3]Department of Music, University of York, UK

In contrast to their traditional, acoustic counterparts, digital musical instruments (DMIs) rarely feature a clear, causal relationship between the performer's actions and the sounds produced. Instead, they often function simply as controllers, triggering sounds that are synthesised elsewhere. Consequently, the performer's interaction with the device frequently does not appear to correlate directly with the sonic output, making it difficult for spectators to discern how gestures are translated into sounds. The input-output relationship is determined by the mapping, the term for establishing causal connections between the control and sound generation parts of the DMI. This study from the 3DMIN (Design, Development and Dissemination of New Musical Instruments) project aims to shed light on how the level of perceived causality of mapping designs impacts spectator responses to new DMIs. Thirty-one participants rated video clips of performances with DMIs with causal and acausal mapping designs in their original version and in a manipulated version. It was predicted that the causal DMIs would be rated more positively, with the manipulation having a stronger effect on the ratings for the causal DMIs. The results indicated that a lack of perceptible causality does have a negative impact on ratings of DMI performances. The instruments in the causal group were viewed as considerably more interesting and more successful at holding the participants' attention than the acausal instruments and in comparison to their manipulated versions. The acausal group received no significant difference in ratings between original and manipulated clips. We posit that this result arises from the greater understanding that clearer gesture-sound causality offers spectators. The ability to perceive reliable relationships between gesture and sound establishes a necessary foundation for further judgements of a performance. The implications of this result for DMI practice will be discussed.

Do Melodic Intervals Evoke Distinctive Qualia?

Lindsay Warrenburg[1] and David Huron[1]
[1]School of Music, Ohio State University, Columbus, Ohio, USA

Individual scale tones are known to evoke stable qualia for experienced listeners (Arthur, 2015; Huron, 2006). In this study, we ask whether melodic intervals evoke distinctive qualia, and whether these qualia are independent of the scale degree qualia of the constituent tones. The first study was an open-ended exploratory task in which musician listeners provided free association terms in response to ascending and descending melodic intervals in both tonal and atonal contexts. Content analysis was carried out on the resulting terms using independent assessors. In order to test the validity of the qualia dimensions, a second study asked independent participants to judge melodic intervals according to the dimensions arising from the content analysis from the first study. In particular, we tested for intersubjective reliability suggesting that interval qualia are stable experiences shared by Western-enculturated listeners. Even if the dimensions are shown to exhibit high reliability, a possible objection would be that listeners are actually responding to the qualia evoked by the constituent scale degrees rather than by the intervals per se. Accordingly, a third study was conducted in which participants again heard melodic intervals in both key contexts and atonal contexts. Intervals were judged using only those dimensions validated in Study 2. Two analyses were used to interpret the results of Study 3. First, do the qualia judgments for key-contextualized intervals converge with the qualia judgments for the a-contextualized intervals? That is, is there any evidence that intervals have stable qualia independent of key context? Second, can we predict the qualia of the contextualized intervals as some combination of the qualia of the constituent tones? For example, if the qualia of an interval is similar to the qualia of the final tone forming that interval, then the idea that intervals evoke independent qualia would be thrown into question.

The Constituents of Art-Induced Pleasure — A Critical Integrative Literature Review

M. Tiihonen[1], E. Brattico[2], and S. Saarikallio[1]
[1]Finnish Centre for interdisciplinary Music Research, Department of Music, University of Jyväskylä, Jyväskylä, Finland
[2]Center for Music in the Brain (MIB), Department of Clinical Medicine, Aarhus University & Royal Academy of Music Aarhus/Aalborg, Denmark

In order to provide a starting point for investigating aesthetic experiences induced by music and visual art in the same framework, the present literature review compares how pleasure has been conceptualized in empirical music and visual art research over the past twenty years. Using the keywords of pleasure, reward, enjoyment, and hedonic seven databases were searched. Twenty music and eleven visual art papers were systematically compared to answer the following questions: 1. What is the role of the keyword in the research question? 2. Is pleasure considered a result of variation in the perceiver's internal or external attributes? 3. What are the most commonly employed methods and main variables in empirical settings? The results demonstrate that musical pleasure is often measured as neural activation in the reward brain circuit combined with subjective ratings of pleasantness or valence. Conceptually, music studies often target different emotions, their intensity and lack of reward, and their relation to the personality and background of the perceiver. Visual art pleasure is also frequently addressed using brain imaging methods, but the focus is on sensory cortices rather than the reward circuit alone. Visual art research more frequently investigates pleasure in relation to cognitive processing and aesthetic experiencing, and the variations of stimulus features and the viewing mode are regarded as explanatory factors of the derived pleasure or aesthetic experience. We propose an integration of the two fields toward a more comprehensive understanding of pleasure, in which intrinsic pleasure is not mutually exclusive of higher-order pleasure and the perceiver is acknowledged to play both passive and active roles in the experience. Additionally, rather than considering pleasure as a static response, scholars should consider how art-induced pleasure develops and fluctuates temporally while interacting with other behavioral, cognitive and affective processes.

"Leaky Features": Influence of the Main on the Many

Michael Barone[1], Matthew Woolhouse[2]
[1] *Psychology, Neuroscience, & Behaviour, McMaster University, Canada*
[2] *School of the Arts, and the McMaster Institute for Music and the Mind, McMaster University, Canada*

The digitization and streaming of music online has resulted in unprecedented access to high quantities of music-information and user-consumption data. Publicly available resources such as Echo Nest have analyzed, and extracted audio features for millions of songs. We investigate whether preferences for audio features in a listener's main genre are expressed in the other genres they download, a phenomenon we refer to as "leaky features". For example, do people who prefer Metal listen to faster and louder Classical or Country music (Metal typically has a higher tempo than many other genres)? A music-consumption database, consisting of over 1.3 billion downloads from 17 million users between 2007-14, is used to explore 10 high level audio features, grouped into 3 categories: rhythmic (danceability, tempo, duration), environmental (liveness, acousticness, instrumentalness), and psychological (valence, energy, speechiness, loudness). The study was made possible with a data-sharing and cooperation agreement between McMaster University and Nokia Music, a global digital music streaming service. Genre information was provided directly by the labels. Audio features for 7 million songs were collected using Echo Nest API, and linked to the Nokia Music database. Although not exhaustive, audio features utilized in this study represent a core set of objective values by which music audio is currently analyzed. We found audio features specific to users' main genres "leaked" into other downloaded genres. Audio features that tend to be more exaggerated in one genre, such as "speechiness" in Rap music, typically demonstrate more leakiness than genres without this characteristic, such as Classical. These results have implications for music-recommendation systems, and future music cognition studies. Additionally, this work demonstrates a unique approach to conducting psychological research exploring the extent to which musical features remain tied to specific genres.

The Silent Rhythm and Aesthetic Pleasure of Poetry Reading

Natalie M. Phillips[1], Lauren Amick[2], Lana Grasser[1], Cody Mejeur[1], J. Devin McAuley[1,3]
[1]Digital Humanities and Literary Cognition Lab, Michigan State University, East Lansing, MI, USA
[2]Department of Psychology, Michigan State University, East Lansing, MI, USA
[3]Neuroscience Program, Michigan State University, East Lansing, MI, USA

In the past decade, there has been increasing interest in the relation between music and language processing. One area of overlapping research is in the domain of rhythm. In this study, we examined individual differences in aesthetic responses during silent reading of sonnets – a highly structured literary form in iambic pentameter that shares a number of rhythmic characteristics with music. Participants read sixteen sonnets at their own self-defined pace, first for familiarity and then to make judgments about different characteristics of the poems. For the second reading of each sonnet, participants judged vividness of the poem's imagery, valence of poem theme, valence of personal feelings about the poem, strength/intensity of their feelings, and aesthetic pleasure (AP). To examine individual differences in poetry perceptions, we also assessed, prior to the first reading, both positive and negative affect using the PANAS, and considered the role of visual and auditory imagery abilities using the VVIQ and BAIS survey measures, respectively. We also considered potential differences associated with music and literary training. As an online measure of AP, literary trained participants highlighted words within the sonnets that they felt were particularly aesthetically pleasing (positive) or displeasing (negative). Results revealed that strength/intensity of feelings, valence of feelings, and positive affect accounted for approximately 76% of the variance in AP ratings. Moreover, we found that positive highlighting was positively correlated with AP ratings; to our surprise, this contrasted with negative highlighting, which had no impact on AP ratings. Although neither imagery vividness nor visual/auditory imagery abilities contributed substantially to AP, vividness was a greater predictor of AP for individuals with high imagery abilities. Findings will be discussed in the context of the relation between rhythm and aesthetic pleasure in music and language more broadly.

Visual Imagery and the Semantics of Music

Paige Tavernaro[1], Elizabeth Hellmuth Margulis[1]
[1]Department of Music, University of Arkansas, Fayetteville, AR, USA

Music triggers visual imagery quite easily (Osborne, 1980). This imagery influences the meaning people ascribe to music and the emotions they experience while listening (Juslin & Västfjäll, 2008). The musical structure, its contextual associations, and characteristics of the listener likely all play a role in stimulating visual imagery. This study aims to understand how visual imagery responses are affected by extramusical descriptions, and what role individual differences play. One hundred forty participants were randomly assigned to one of three groups: congruent, incongruent, or no description. Each group heard the same nine orchestral excerpts in a randomized order. A prior study with twenty-four participants established a set of brief descriptions of visual imagery that seemed congruent or incongruent with each excerpt. Participants in the congruent group and incongruent group read a description of congruent or incongruent visual imagery, respectively, before each excerpt. Participants in the third group listened without an imagery prompt. After each excerpt, participants answered questions about their experience of the music. At the end of the study, participants completed the Goldsmiths-MSI and the Absorption in Music Scale. Linear mixed models were conducted with the group as the independent variable and imagery experience, emotional reaction to music, enjoyment of music, and engagement with music as the dependent variables. Results revealed that incongruent descriptions had a negative impact on participants' visual imagery to the music. The congruency or incongruency of the prompts also had interesting effects on the free description reports of visual imagery. Additionally, absorption in music was found to correlate positively with the overall imagery experience of the listener. This study lays the foundation for future work on visual imagery to music.

Disliked music: The other side of musical taste

Taren Ackermann[1]
[1]Music Department, Max Planck Institute for Empirical Aesthetics, Frankfurt, Germany

Musical taste, understood as an attitude towards music, plays an important part in the way music listeners in Western cultures perceive and construct their self-concept. Listeners like or dislike specific music not only to satisfy their emotional and communicative needs, but also to create and affirm their own identity. But until now, research has focused mainly on the positive aspects of musical taste. To get to a better understanding of it, qualitative in-depth-interviews (N = 21) were conducted to explore the different dimensions and justifications individual participants offered in respect to music they particularly dislike. Prior to interview sessions the participants were asked to list their musical dislikes. Then, during the interviews they were asked to give reasons for why they dislike each musical piece, artist or style given on their list and rate each item on a 10-point scale (0=neutral to 10=worst possible item). All interviews were analyzed using qualitative content analysis. The results from this analysis show that reasons for likes and dislikes do not necessarily form opposite pairs, but are in some instances identical. Liking and disliking insofar serves the same function that listeners use both to express their identity and encourage social contact and cohesion. But disliking certain music plays a prominent role when listeners want to avoid negative emotional states and moods, physical harm or unpleasant social situations. Also, all participants confirmed that they can draw links between individual dislikes and their own identity. Ultimately, the study provides further insight into the dimensions and functions of musical taste in general and disliked music in particular, and into how specifically individual dislikes are relevant for the creation and affirmation of one's self-concept.

On the asymmetry of cross-modal effects between music and visual stimuli

Teruo Yamasaki[1]
[1]Department of Psychology, Osaka Shoin Women's University, Osaka, Japan

There are two lines of findings on the cross-modal effects between music and visual stimuli. One claims that the effect of music on the impression of visual stimuli dominated the effect of visual stimuli on the impression of music, and the other does vice versa. In the author's previous study, the cross-modal effect of music dominated one of paintings with relatively long exposure, while this dominance of music disappeared with brief exposure. As one possibility, the steadiness of stimuli could explain these results. Music changes continuously but paintings don't. With long exposure, the effect of music accumulates, whereas the effect of paintings may not grow because of habituation. This study examined the influence of steadiness on the cross-modal effects. In order to control the steadiness, an affective picture was presented with 20 seconds in the single picture condition and 20 pictures were presented in succession with a second each in the 20 pictures condition. Music excerpts were 4 piano solos. Participants rated the potency and brightness impression of music and pictures individually in the first part of the experiment. In the second part, participants were presented music and pictures at the same time and rated the impression of music or pictures. To compare the cross-modal effects between music and pictures, musical impact and pictorial impact were calculated. Musical impact was defined as the difference between the potency (or brightness) scores of a visual stimulus with and without a musical excerpt. Pictorial impact was also defined in the same way. As the results, for potency, ANOVA revealed that the musical impact was significantly larger than the pictorial one in the single picture condition. On the other hand, in the 20 pictures condition, such a significant difference was not found. For brightness, similar tendency was shown. These results suggest that the steadiness of stimulus influences the cross-modal effects between music and visual stimuli.

Zoning in or Tuning in? Absorption within a Framework for Attentional Phases in Music Listening

Thijs Vroegh[1]
[1]Music Department, Max Plank Institute for empirical Aesthetics, Frankfurt am Main, Germany

In terms of theoretical and empirical initiatives, Herbert (2011) put music absorption back on the map of scholarly interests. However, having focused exclusively on everyday life experiences, she and others have left absorbed listening in settings that generally seem to require and assume full attention largely unexplored. The main aim of this study was to gain insight in the underlying structure of 'aesthetic absorption' while listening to music. A second aim was to distinguish between possibly different types of absorption which may occur as part of aesthetic experiences, and to understand them in relation to other types of attentional behavior when listening to music. Via an online survey (N=228) participants were asked to listen to a self-chosen piece of highly-involving music. Data were gathered on a variety of parameters related to consciousness and attention. Factor analysis (PAF) and the Schmid-Leiman procedure provided a deeper and parsimonious understanding in the higher-order structure of consciousness while being musically engaged. Its cognitive-motivational part was used to further define the dimensionality of absorption, and was linked to Tellegen's model of absorption concerning the experiential and instrumental sets. Next, a hierarchical cluster analysis was conducted in order to explore different types of attentional behavior. Results suggested distinguishing between two types of musical absorption which differ in the presence of meta-awareness. These were termed 'zoning-in' and 'tuning-in' (cf. Smallwood 2007). Finally, a dynamic framework was developed which integrates these types with other attentional phases (e.g., mind wandering) applicable to music listening. It is argued that this framework may be a realistic alternative to the often problematic distinction between everyday aesthetics and art-centered aesthetic experiences.

Effects of the duration and the frequency of temporal gaps on the subjective distortedness of music fragments

Kunito Iida[1], Yoshitaka Nakajima[2], Kazuo Ueda[2], Gerard B. Remijn[2], Yukihiro Serizawa[3]
[1]Graduate School of Design, Kyushu University, Fukuoka, Japan
[2]Department of Human Science/Research Center for Applied Perceptual Science, Kyushu University, Fukuoka, Japan
[3]Automobile R&D Center, Honda R&D Co., Ltd., Tochigi, Japan

The sound quality of car radios is often degraded by poor radio-wave reception. Ways to estimate the qualities of sounds transmitted by radio are sought after. We attempted to establish a psychophysical way to evaluate the subjective distortedness of music in which some parts were interrupted, as is often the case in car radio sound. A magnitude estimation experiment was conducted. Two music sources of about 5 s, classic and rock, were degraded by replacing some parts with temporal gaps. The number of gaps per second was 2-16, and the duration of each gap was 6-48 ms, with 17 stimuli for each music source. Twelve participants estimated the subjective distortedness of each stimulus, responding with a number above or equal to zero. Zero meant that no distortion was perceived. The results showed that magnitude estimation turned out to be a reliable method to measure the subjective distortedness of music degraded by temporal gaps: the subjective distortedness was described as a power function (with exponent of about 0.7) of gap duration summed up per second.

Modeling Audiovisual Tension

Morwaread M. Farbood

Dept. of Music and Performing Arts Professions, New York University, USA

mfarbood@nyu.edu

ABSTRACT

An experiment was conducted to examine the perception of audiovisual tension. Subjects provided real-time tension judgments of 2'30" stimuli in one of three categories: visual animation with audio, visuals alone, or audio alone. The animations consisted of random-dot kinematograms and the audio consisted of three excerpts taken from electronic art music. The animations were described by changes in the features that were used to generate them: contrast, speed of motion, coherence of motion, and brightness. The musical excerpts were described in terms of loudness, pitch, onset frequency, and the timbre attributes inharmonicity, roughness, spectral centroid, spectral deviation, and spectral flatness. Initial analysis consisted of correlations between auditory/visual features and mean subject responses. A computational model based on trend salience was then used to predict subject judgments of tension given the visual and auditory descriptors. The model was able to predict judgments well, resulting in high correlations between predictions and mean tension responses.

I. INTRODUCTION

While many aspects of auditory and musical tension have been explored in prior work, there has been a relative dearth of empirical work on timbral features contributing to perception of tension and no work that has examined timbral tension in combination with visual tension. Based on prior work on timbre (Bailes & Dean, 2012; Dean & Bailes, 2010; Farbood & Price, 2014; Pressnitzer, McAdams, Winsberg, & Fineberg et al., 2000, Schubert, 2004) and methodological approaches in vision research (cf. Baker & Braddick, 1982; Schütz, Braun, Movshon, & Gegenfurtner, 2010), several auditory and visual features were used to model tension responses to audiovisual stimuli.

The purpose of the study was to examine contributions of specific auditory and visual features to audiovisual tension. The auditory features included the five timbre descriptors inharmonicity, roughness, spectral centroid, spectral deviation, spectral flatness as well as loudness, pitch, and onset frequency. The visual features included speed of motion, coherence of motion, visual contrast, and visual brightness. The goal was to better understand the relative contributions of these features and to empirically determine the weights of their individual contributions to perceived tension.

II. METHOD

An experiment was conducted in which 45 participants were asked to judge how they felt tension was changing by moving a slider while observing and listening to 15 audiovisual stimuli. Each stimulus paired one of four audio files with one of four visual animations. The audio consisted of three excerpts from electronic compositions by Nono, Stockhausen, and T. H. Park, or silence, and the visual animations consisted of different types of changes in the visual parameters, or no visuals (static black screen). The animations were designed to reflect three general types of changes: random changes, gradual ramps, and shorter ramps. These changes were visualized using a monochromatic random-dot kinematogram. The number and size of dots were generated to take up exactly half of the available screen space in order to allow for equal proportions of foreground and background brightness.

III. RESULTS

The first step in the analysis was to examine how the different auditory and visual features correlated with mean subject responses. Table 1 shows the correlation coefficients (Spearman's rho) for each of the 15 stimuli. Although the *r*-values vary considerably between stimuli, the means for each feature provide an initial suggestion for which features might significantly influence perceived tension; in particular, these include loudness, roughness, and onset frequency among the auditory features, and speed of motion for the visual features.

In order to better understand the contribution of the various features to tension judgments, a trend-salience model, developed for musical tension (Farbood, 2012), was used to predict the empirical data. The model integrates tension contributions from individual features by taking into account the cumulative slope of those features within a moving "attentional" window and adjusting the slope based on whether it is a directional continuation of what happened immediately before it (the "memory" window). The attentional window snapshots are then merged as overlapping windows in time.

The durations of the memory and attentional windows as well as the memory window weight are all variables and that can be adjusted to improve the predictive power of the model. The values used in this case were mostly derived from prior work (Farbood, 2012). However, deciding the feature weights (and whether they were even necessary) were obtained in the current study through a manner similar to stepwise regression: all features were given equal weight initially and then individually removed to see if the predictions improved. The decision to retain or eliminate features and assign relative weights were done using the audio-only and visual-only stimuli.

The optimized model for the audio-only stimuli retained loudness, spectral centroid, roughness, and onset frequency. All weights were equal, except for loudness, which was twice that of the other features. The mean correlation coefficient (Spearman's rho) between the audio-only stimuli and the predictions produced by the model was .81. The optimized model for the visual-only stimuli retained all four visual features—speed of motion, contrast, coherence, and brightness (respective weights 6, 3, -1, -1)—with speed having the greatest weight followed by contrast. The mean correlation coefficient value (Spearman's rho) between the visual-only stimuli and the predictions was .72.

Table 1. Correlations between features and responses for each stimulus. Abbreviations: Aud = musical excerpt; Vis = animation; Lou = loudness; Cen = spectral centroid; Dev = spectral deviation; Inh = inharmonicity; Rou = roughness; Fla = spectral flatness; Pit = pitch; Ons = onset frequency; Spe = speed of motion; Coh = coherence of motion; Con = contrast; Bri = brightness.

Stimulus		Features											
		Audio								Visual			
Aud	Vis	Lou	Cen	Dev	Inh	Rou	Fla	Pit	Ons	Spe	Coh	Con	Bri
A1	-	.71	.04	.20	.02	.21	.29	-.01	.68	-	-	-	-
A2	-	.53	-.44	-.63	.04	.58	-.61	.28	.57	-	-	-	-
A3	-	.69	.65	.63	.40	.28	.53	.18	.12	-	-	-	-
A1	V1	.55	.17	.32	.09	-.01	.41	-.10	.69	.26	.20	.42	-.33
A1	V2	.71	.10	.24	-.04	.10	.34	.02	.73	.41	.39	-.06	.22
A1	V3	.75	.10	.24	-.01	.21	.26	.04	.62	.26	.17	-.28	-.23
A2	V1	.22	-.22	-.25	.07	.30	-.28	.39	.44	.37	-.16	.27	-.28
A2	V2	.30	-.31	-.38	.10	.38	-.40	.40	.48	.56	.52	.03	-.07
A2	V3	.36	-.42	-.42	.03	.33	-.43	.25	.46	.42	.22	-.32	-.14
A3	V1	.41	.32	.23	.21	.39	.20	.23	.17	.33	-.21	.04	-.03
A3	V2	.26	.22	.23	.17	.12	.18	.17	.36	.41	-.16	.30	-.42
A3	V3	.37	.35	.39	.31	.18	.29	.09	-.08	.39	.06	.17	.08
-	V1	-	-	-	-	-	-	-	-	.38	-.22	.39	.06
-	V2	-	-	-	-	-	-	-	-	.58	.29	.39	-.13
-	V3	-	-	-	-	-	-	-	-	.59	-.06	-.16	-.30
Mean		.49	.05	.07	.12	.26	.07	.16	.44	.41	.09	.10	-.13

When tested on stimuli with both audio and visual components, the best results were obtained when audio features were given twice the weight (or more) of the visual features. The optimized model for the audio + visual stimuli resulted in a mean correlation value of .67 (min of .31, max of .87). In general the model did quite well, particularly in predicting more local tension changes.

IV. CONCLUSION

This paper reports a preliminary analysis of an experiment that explored the perception of audiovisual tension. Subjects were asked to judge perceived tension when watching/listening to stimuli featuring three musical excerpts taken from electronic compositions paired with random-dot animations. A trend-salience model was utilized as a way of determining which auditory and visual features were influencing tension perception and the relative contributions of those features. When the model was optimized to fit the mean responses, the feature with the greatest weight was found to be loudness, followed by spectral centroid, roughness, onset frequency, and speed of motion (all equal in weight), then by contrast, and finally with much smaller negative contributions from coherence of motion and brightness. In general, the auditory features contributed significantly more to perceived tension than the visual features.

REFERENCES

Bailes, F., & Dean, R. T. (2012). Comparative time series analysis of perceptual responses to electroacoustic music. Music Perception, 29(4), 359–375.

Baker, C. L., & Braddick, O. J. (1982). The basis of area and dot number effects in random dot motion perception. Vision Research, 22(10), 1253–1259.

Dean, R. T., & Bailes, F. (2010). Time series analysis as a method to examine acoustical influences on real-time perception of music. *Empirical Musicology Review, 5*(4), 152–175.

Farbood, M. M. (2012). A parametric, temporal model of musical tension. Music Perception, 29(4), 387–428.

Farbood, M. M., & Price, K. (2014). Timbral features contributing to perceived auditory and musical tension. In *Proceedings of the 13th International Conference of Music Perception and Cognition*, Seoul, Korea.

Paraskeva, S., & McAdams, S. (1997). Influence of timbre, presence/absence of tonal hierarchy and musical training on the perception of musical tension and relaxation schemas. In *Proceedings of the 1997 International Computer Music Conference*, 438-441.

Pressnitzer, D., McAdams, S., Winsberg, S., & Fineberg, J. (2000). Perception of music tension for nontonal orchestral timbres and its relation to psychoacoustic roughness. *Perception & Psychophysics, 62*(1), 66–80.

Schubert, E. (2004). Modeling perceived emotion with continuous musical features. Music Perception, 21(4), 561–585.

Schütz, A. C., Braun, D. I., Movshon, J. A., & Gegenfurtner, K. R. (2010). Does the noise matter? Effects of different kinematogram types on smooth pursuit eye movements and perception. *Journal of Vision, 10*(13), 26–26.

Computational Modeling of Chord Progressions in Popular Music

Stefanie Acevedo

Department of Music, Yale University, USA
stefanie.acevedo@yale.edu

I. BACKGROUND AND AIMS

Many computational analyses of popular music harmony focus on probabilistic modelling of unigrams and chord-to-chord transitions (Burgoyne, 2011; de Clercq & Temperley, 2011; White & Quinn, in press). This research often serves to model functional states of individual chords, especially in comparison to common-practice tonal functions (Tonic, Subdominant, Dominant). However, due to their bottom-up focus, these analyses are often divorced from larger contextual parameters, such as meter and formal structure, which some scholars argue are the driving rhetorical characteristics of popular music (Moore, 2001; Tagg, 2014). The aim of this computational study is to identify fundamental harmonic progressions, including variants (based on multiple chord characteristics such as chord root or quality), within a popular music corpus, and also to identify and categorize these progressions based on contextual parameters including metrical placement and phrase/formal structure.

II. METHODS

Using a Python script, *n*-gram and transition probabilities are calculated within a sample taken from Billboard's *Hot 100* songs from 1958 to 1991 (the McGill Corpus; Burgoyne, Wild, & Fujinaga, 2011; Burgoyne, 2011). The program then segments each song into end-bounded chord progressions based on entropy spikes between a stipulated progression of chords (the prefix) and a succeeding chord (the suffix). This is done by iterating through each possible *n*-gram of a song and testing that: A) the entropy of the suffix is greater than a given threshold ($H > 0.9$); and B) the probability of the last prefix element is higher than a given threshold ($P > 0.5$). Given these two parameters, the prefix is encoded as a progression, each of which is tabulated based on metrical and formal information.

III. RESULTS

The most common chord sequences include repetitions of two-chord loops (Tagg, 2014) and three-, four-, or five-chord stock progressions (such as the *Doo Wop*, and modal or pentatonic loops; Everett, 2009; Biamonte, 2010). Based on number of repetitions, the metrical placement of specific chords, formal location, and ending chords, the progressions are classified by apparent function (for example, as embellishments, prolongations, or cadential gestures). Further analysis of chord-functional substitutions, as stipulated in previous popular music unigram models (White & Quinn, in press), allow for categorization of variant sequences into broader families of progressions.

IV. SUMMARY AND FUTURE DIRECTIONS

The results are discussed within two larger contexts: that of probabilistic mental models (couched within current discourse on musical schemata and expectation research; Byros, 2012; Gjerdingen, 2006) and their relationship to broad and genre-specific definitions of *tonality*. The results of this research will be employed for ERP experimentation to supplement current research on functional and harmonic expectation (Koelsch, Rohrmeier, Torrecuso, & Jentschke, 2013; Koelsch, 2009; Ruiz, Koelsch, & Bhattacharya, 2009).

ACKNOWLEDGMENT

The author thanks her dissertation advisor, Ian Quinn, for support and advice, as well as Christopher Wm. White and David Temperley for analytical guidance. Funding for this project comes from a Seed Grant provided by Yale University's Digital Humanities Lab.

REFERENCES

Biamonte, N. (2010). Triadic modal and pentatonic patterns in rock music. *Music Theory Spectrum*, *32*(2), 95-110.

Burgoyne, J. A. (2011). *Stochastic Processes & Database-Driven Musicology*. PhD Dissertation, McGill University.

Burgoyne, J. A., Wild, J., & Fujinaga, I. (2011). An Expert ground-truth set for audio chord recognition and music analysis. *12th International Society for Music Information Retrieval Conference* (ISMIR), 633–638.

Byros, V. (2012). Meyer's Anvil: Revisiting the schema concept. *Music Analysis*, *31*, 273–346.

de Clercq, T., & Temperley, D. (2011). A corpus analysis of rock harmony. *Popular Music*, *30*(1), 47–70.

Everett, W. (2009). *The Foundations of Rock*. New York: Oxford University Press.

Gjerdingen, R. (2006). *Music in the Galant Style*. New York: Oxford University Press.

Koelsch, S. (2009). Neural substrates of processing syntax and semantics in music. *Music That Works: Contributions of Biology, Neurophysiology, Psychology, Sociology, Medicine and Musicology*, 143–153.

Koelsch, S., Rohrmeier, M., Torrecuso, R., & Jentschke, S. (2013). Processing of hierarchical syntactic structure in music. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(38), 15443–8.

Moore, A.F. (2001). *Rock: The Primary Text*. Aldershot: Ashgate.

Ruiz, M. H., Koelsch, S., & Bhattacharya, J. (2009). Decrease in early right alpha band phase synchronization and late gamma band oscillations in processing syntax in music. *Human Brain Mapping*, *30*(4), 1207–1225.

Tagg, P. (2014). *Everyday Tonality II. Towards a Tonal Theory of What Most People Hear*. New York, Huddersfield: The Mass Media Music Scholars' Press.

White, C. W., & Quinn, I. (in press). Some Empirical Models of Harmonic Function. *Music Theory Spectrum*.

Beneath (or beyond) the surface: Corpus studies of tonal harmony

David Sears[1], Claire Arthur[2]
[1]Centre for Research in Music Media and Technology (CIRMMT), Schulich School of Music, McGill University, Montreal, QC, CANADA
[2]School of Music, Ohio State University, Columbus, OH, USA

For over two centuries, scholars have observed that tonal harmony, like language, is characterized by the logical ordering of successive events, what has commonly been called *harmonic syntax*. Yet despite the considerable attention devoted to characterizing the behavior of chord events at relatively local levels of musical organization, the application of statistical modeling procedures—advanced in the cognitive sciences to explain how humans learn and mentally represent languages—has yet to gain sufficient traction in music research.

At present, few harmony corpora exist in the scholarly community, and many consist of symbolic representations of homo-rhythmic genres that rarely feature in everyday listening. Perhaps worse, corpus studies often suffer from the *contiguity fallacy*, the assumption that note or chord events on the musical surface depend only on their immediate neighbors, despite the wealth of counter evidence from studies of sequence memory that listeners form non-contiguous associations for stimuli demonstrating hierarchical structure.

This session presents four corpus studies of tonal harmony that look beneath (or beyond) the surface, examining problems relating to the representation and analysis of tonal materials using techniques for pattern discovery, classification, and style analysis. John Ashley Burgoyne (The Netherlands) examines the evolution of expected per-song chord distributions over the period 1958–1991, based on the *McGill Billboard Project* (http://billboard.music.mcgill.ca/), an open-access corpus of expert harmonic transcriptions for over 1,000 songs from the Billboard "Hot" 100. David Sears (Canada) demonstrates the utility of non-contiguous *n*-grams for the discovery of recurrent harmonic patterns in a corpus of Haydn's string quartets, where non-contiguous events often serve as focal points in the syntax. Christopher White (USA) presents an algorithm that derives a reduced chord vocabulary from the surface statistics in the *Yale Classical-Archives Corpus* (http://ycac.yale.edu/) without recourse to pre-made harmonic templates or knowledge of tonal harmony in general, with the resulting vocabulary and syntax conforming to those produced by humans. Finally, Claire Arthur (USA) considers whether algorithms based only on melodic features—metric position, duration, and interval of approach and departure—can identify *non-chord tones* in classical and popular corpora, using harmonic transcriptions as a basis for comparison.

‘Revolutions’ in pop music: The experts vs. the audio

John Ashley Burgoyne[1]
[1]Institute for Logic, Language and Computation, University of Amsterdam, The NETHERLANDS

Using MIR techniques on an audio corpus, Mauch et al. (2015) identified three moments of change in Western popular music style: the years 1964, 1983, and 1991. As yet, their claims have not been tested against human transcriptions of the musical material. We are seeking to verify and enrich these findings with the McGill *Billboard* corpus (http://billboard.music.mcgill.ca/), a collection of over 1000 expert transcriptions of the harmony and formal structure of songs from the U.S. pop charts.

Symbolic corpus analyses are bedeviled by two common challenges: (1) correctly handling the sum-to-one constraint implicit in considering relative frequencies of chords within pieces, composers, or styles; and (2) measuring evolution smoothly over time, allowing inflection points to appear naturally, without introducing arbitrary cut-points. We converted the relative frequencies of chord roots and bigrams to so-called “balances” using the principles of compositional data analysis (Burgoyne et al., 2013). This approach avoids spurious correlations that can appear with techniques that ignore the sum-to-one constraint. We then analyzed the evolution of these balances using time-series analysis, rather than grouping according to fixed time spans.

The data do not support a strong inflection point in 1964, though there is a small reduction in the usage of vi chords and descending minor third progressions. There is more support for 1983, where the use of bVII chords, minor keys, and stepwise progressions strongly peak. The corpus ends in 1991, making it impossible to verify the third potential inflection point. Together, these results are consistent with Mauch et al.'s audio findings: a small stylistic change in the mid-1960s, followed by smooth and steady evolution until a sharp break occurred in the early 1980s. Principal component analysis further reveals that the variation within harmonic vocabulary is substantially more complex than that of the root progressions themselves.

A corpus study of tonal harmony: Pattern discovery using non-contiguous *n*-grams

David Sears[1], Stephen McAdams[1]
[1]Centre for Research in Music Media and Technology (CIRMMT), Schulich School of Music, McGill University, Montreal, QC, CANADA

Pattern discovery is an essential task in many fields, but particularly so in the cognitive sciences, where a substantial body of evidence suggests that humans are more likely to learn and remember the patterns they encounter most often. In corpus linguistics, researchers often discover recurrent patterns by dividing the corpus into contiguous sequences of *n* events, called *n*-grams, and then determining the frequency of each distinct pattern in the corpus. And yet for stimuli demonstrating hierarchical structure, non-contiguous events often serve as focal points in the syntax. This problem is particularly acute for corpus studies of tonal harmony, where the musical surface contains considerable repetition, and many of the vertical sonorities from the notated score do not represent triads or seventh chords, thereby obscuring the most recurrent patterns.

We present an alternative approach to pattern discovery by also including non-contiguous *n*-grams, one that rekindles Hermann Ebbinghaus' *doctrine of remote associations*, in which associations are strongest between contiguous events and become progressively weaker for increasingly non-contiguous events. The corpus consists of combinations of chromatic scale degrees measured at each unique onset in 50 sonata-form expositions from Haydn's string quartets. To account for the decay in sensory or working memory resulting from the potential discontiguity between adjacent members of a given pattern, we weight the appearance of each *n*-gram using an inverse exponential function with a half-life of 1 s such that patterns further separated in time receive smaller weights in the final count. Our results demonstrate that for contiguous *n*-grams, the power-law relationship between frequency and rank disappears as the value of *n* increases, but this relationship returns for frequency tables based on non-contiguous *n*-grams. Moreover, the most conventional harmonic progressions emerge as among the most frequent patterns in the corpus.

Deriving and Evaluating SPOKE, a Set-Based Probabilistic Key Finder

Christopher William White*[1]

*Department of Music and Dance, The University of Massachusetts, Amherst, USA
[1]cwmwhite@umass.edu

ABSTRACT

When listeners attribute a key to a passage of music, they are cognizing pitch events as scale degrees, or members of a scale in some key. This task is likely passively learned through exposure to the statistical properties of tonal music, allowing a listener to infer a pitch's scale degree identity based on some observable property. Much computational modeling of the learning and execution of this task has focused on pitch events, with the models' parameters being derived from corpora and the models' behaviors corresponding to human key assessments. This study presents an alternate chord-based model that both learns its parameters from a corpus, and also tracks human behavior. The engineering behind the learning task is described, relying on contextual regularities and subset/superset relationships to produce a chordal vocabulary and syntax approximating those published in several music-theory textbooks. Three tests follow, each designed to evaluate the model's relationship to human behavior. The results show that the chord-progression model exhibits precision approximating human performance and even outperforms certain established pitch-based key-finding models.

I. INTRODUCTION

Figure 1 shows the opening passage of Mozart K. 284, iii. Viewing the music inside the two boxes, a musician would likely hear the first passage in D major, with the second modulating to A major.

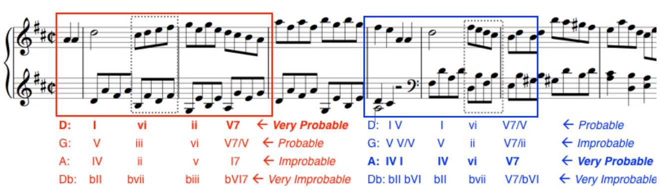


Figure 1. A toy analysis of Mozart K. 284, iii, mm. 1-8

How do we undertake this key finding task? Much emphasis in music-perception research has focused on the distribution of pitches in tonal contexts to answer this question, comparing the relative amount of onsets or durations of each pitch to some ideal template. For instance, listeners generally associate the most frequent tone with the first degree of the major or minor scale, the second most frequent tone with scale-degree five, and so on (Krumhansl & Shepard, 1979; Castellano, Bharucha, & Krumhansl, 1984; Oram & Cuddy, 1995; Creel & Newport, 2002; Smith & Schmuckler, 2004). In Figure 1's initial window, D and A occur most often, thereby aligning this passage with a D major key center. Importantly, this view of tonal induction connects not only to cognition, but also to learning: since the scale-degree distributions that listeners associate with major and minor keys correspond to distributions found in corpora of classical music, such a key-finding model could arise from the statistical learning that might accompany a listener's exposure to such repertoires (Huron, 2006; Temperley, 2007).

But pedagogical practice differs from this paradigm. In the music theory classroom, students are taught to identify a passage's key by matching surface events with successions of key-oriented chord symbols. Most mainstream pedagogy (e.g. Laitz, 2008; Roig-Francoli, 2010; Aldwell & Schachter, 2011) focuses on the major and minor triads rooted on each scale degree acting as the basic vocabulary of possible chords – generally shown by Roman Numerals that correspond to the root's scale degree – along with an accompanying syntax that defines how chords tend to progress to one another. Figure 2 provides a heuristic approximation of such a textbook vocabulary and syntax (roughly following Kostka & Payne 2012). Unlike methods that rely on pitch distributions, this method chooses a key by aligning chord symbols in ways that make syntactic sense. In the Roman numerals shown in Figure 1, the D major and A major analyses are chosen as "most probable" because – according to Figure 2 – these sequences of Roman numerals are the most likely to occur.

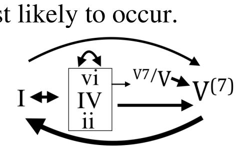


Figure 2: A hypothetical chord progression model, with arrows representing the most probable between-chord successions

However ubiquitous this method might be within pedagogy and practice, few computational models have been implemented that use corpus data to build a key-finding system that relies on chord events (Bharucha, 1987; Hörnel & Ragg, 1996; Barthelemy & Bonardi, 2001; Raphael & Stoddard, 2004; Bellman, 2005; Quinn, 2010; White, 2013), and (to the author's knowledge) none have been evaluated against human behavior. Instead, most key-orientation studies that model event successions have focused on pitch events as well, especially those successions that outline rare intervals that indicate particular key centres (Brown & Butler, 1981; Brown, Butler & Jones 1994; Matsunaga & Abe 2005; 2012).

This paper presents a chordal set-based probabilistic key finder – or, SPOKE – that a) models statistical learning by deriving parameters directly from a corpus, and b) can be compared to human behavior. Such comparisons will then be undertaken in three ways: 1) comparing SPOKE's key assessments to those of a annotated corpus, 2) comparing the model's Roman-numeral analyses to those of undergraduate music majors, and 3) asking professional music theorists to distinguish the computer's output from that of undergraduate music majors. These tests are designed to provide first steps toward entertaining a chordal approach to key finding as a supplement to more established pitch-based tonal modeling.

II. A CHORDAL KEY-FINDING MODEL

A. SPOKE's Key-Finding Method

Models that use chord progressions to find a passage's key translate successions of pitch sets into successions of *scale-degree sets*, or sets of values oriented within a scale (rather than within absolute pitch space). For instance, consider an F-major triad <F-A-C> followed by a C major triad <C-E-G>. Represented as pitch sets, they show no tonal orientation: they could be in almost any key. However, when converted to F-major's scale degrees, the notes are defined by their position within the F-major scale. These two sets become triads on the first and fifth scale degrees, designated respectively as I and V in standard Romans-numeral notation.

The translation between pitch sets and scale-degrees sets has been engineered in several ways, using neural networks (Bharucha, 1987; 1991, Hörnel & Ragg, 1996), Hidden Markov modeling (Raphael & Stoddard 2004), and *n*-gram Markov chains (Quinn 2010, White in press); however, only the last of these has aggressively relied on large corpora to set their models' parameters. This study will therefore rely on this established paradigm. Such a process essentially mimics that used in the Roman-numeral analysis of Figure 1. Imagine the thickness and direction of the arrows of Figure 2 as representing 2-gram probabilities (i.e., the probability of each chord moving to the next chord). From this standpoint, the D major and A major readings are ideal since their Roman numerals produce the highest 2-gram probabilities compared to those from other tonal orientations. It is then a small step to derive these transition probabilities from a corpus by tracking how frequently each chord moves to each other chord within that corpus.

Equation 1 formalizes this process (following White 2014). The equation acts on a progression of events ($o_{j-n} \ldots o_j$), a series of n pitch sets o between two timepoints j-n and j, and finds the key judgment x that transposes that series to scale-degree sets. This key judgment at time point j is the equation's end result, $\hat{x}_j$. More specifically, we use integers [0…11] to represent the initial unanalyzed notes as *pitch classes* (where C is 0, C# is 1, D is 2, B is 11, and so on) and the resulting scale-degree sets as comprised of *scale-degree classes* (where tonic is 0, the leading tone is 11, supertonic is 2, and so on), and apply modulo 12 arithmetic to transpose the former to the latter (Forte 1973). Here, x acts as the example's tonal interpretation, transposing the series to one of the 12 possible tonal centers (i.e., the key of C would be 0, C-sharp would be 1, F would be 5, and so on). To maximize the argument, then, one chooses the key level at which transitions between scale-degree sets produce the highest overall *n*-gram probabilities. In terms of Figure 1, this means that aligning D with tonic (x = 2) would find the ideal key – or $\hat{x}$ – that maximizes the probability of the window – or, argmax(P( ) – at that particular time point j. The contrasting pitch material of the second window's time point changes the probability of the chord sequence: now aligning A with tonic (x = 9) creates the most probable series of sets in this window.

$$\hat{x}_j = \operatorname{argmax}_x(P([o_{j-n}-x]_{\mathrm{mod}\ 12} \ldots [o_j-x]_{\mathrm{mod}\ 12})) \qquad (1)$$
$$x \in |0\ldots11|$$

B. SPOKE's Vocabulary Reduction

But, it is not always obvious which sets should constitute an analytical model's *vocabulary,* or the universe of sets ("chords") possible in the model's tonal analyses. Consider the two dotted boxes of Figure 1. Both boxes have the pitch-class content <C#, D, E, F#, B>. While our toy analysis initially assumed an underlying vocabulary of triads and sevenths, without a priori rules or templates there is no immediate way of culling a B minor triad from that larger set of pitches. To resolve this problem using the properties of the corpus itself, White (2013) proposed a machine-learning *reduction method* for deriving a limited vocabulary of scale-degree sets from a dataset of chord progressions. First, the method observes which progressions of scale-degree sets occur most frequently within some corpus. Those progressions that occur above some frequency threshold comprise the model's initial vocabulary. When presented with some series of scale-degree sets not in this vocabulary, the reduction method adds and subtracts pitches from the observed sets until they are transformed into vocabulary items. As a demonstration of this reduction process, again consider the dotted boxes of Figure 1, imagining that the syntax of Figure 2 represents a corpus's most-frequent progressions. Unreduced, the pitch content of these boxes would not produce chords within the syntax; however, the reduction process would find that, by removing the C# and E, the pitch content of those windows can be transformed into the ii and vi triads, respectively, and therefore participate in syntactic chord progressions.

This process can be iterated over an entire corpus to produce a *reduced vocabulary*: by reducing a corpus's scale-degree transitions, the process "edits" the corpus's less-frequent progressions into their more frequent subsets or supersets (e.g., editing a set like <C#, D, E, F#, B> into <B, D, F#>, or a B minor triad). The process can be repeated – editing the edits – until no more changes are made. Having mapped all improbable sets onto probable ones, the resulting reduced series of chords comprises a vocabulary smaller and more constrained than one culled from the raw musical surface.

Formally, we can represent a series of musical observations O with time points 1 to n such that $O = (o_1, o_2 \dots o_n)$. We can reduce each o in O to a related reduced chord s such that $S = (s_1, s_2 \dots s_n)$ and $|s_i \cap o_i| \geq 1$ where the cardinality of the intersection between each o and its corresponding s is at least 1 (i.e., they share at least one scale-degree). Equation 2 then produces the reduced series S by maximizing the contextual probabilities k and the proximity of the two sets π. Here, $P(s_i|k(o_i))$ is the probability that a given s would occur in the context in which we observe the corresponding o. While this could in principle be any probabilistic relationship, in our case this is a simple bigram probability. $\pi(s_i, o_i)$, then, is the proximity (NB: "π" for "proximity") between the two sets, s and o. While in principle this could be any type of similarity relation, in the current implementation π is equal to the amount of overlap (or intersection cardinality) between the two sets. This method has been shown to produce a vocabulary primarily of triads and seventh chords that conforms to basic intuitions surrounding classical tonal harmony (White 2012). For additional specifics on the implementation, see the study's online supplement, available at chriswmwhite.com/research.

$$S = \text{argmax} \prod_{i=1}^{n} P(s_i|k(o_i))\, \pi(s_i, o_i) \quad (2)$$

SPOKE then combines this reduction process with the key-finding method of Equation 1 to analyze a piece of music. The full process, crucially, determines a musical key by *relying only parameters learned from a corpus*. The model therefore not only produces key-finding assessments that can be measured against human behavior, but it also learns how to undertake this task by replicating exposure-based learning. In what follows, we conduct several tests to evaluate the model's performance as it relates to human behavior. Test 1 examines how accurately the model assesses the key of a ground-truth corpus compared to other established key-finding models.

III. TESTS

C. Test1 – Key-Finding Accuracy

1) Materials and Methods. The reduction process was implemented on the Yale-Classical Archives corpus (YCAC). This corpus is described in White & Quinn (in press) and is available at www.ycac.yale.edu. It consists of 13,769 MIDI files by 571 composers yielding 14,051,144 *salami slices*, or sets of notes each time a pitch is added or subtracted from the texture. Slices are identified with their onset time. The corpus has also been tonally analysed, with sections of each piece identified either with a key and mode or labelled as ambiguous. "Start" and "End" tokens were added to the beginnings and endings of each file. The files were divided into lists of scale-degree sets, and divisions were made at key changes. Each slice was compiled as an unordered set of scale degrees. To further simplify the dataset, singletons and adjacent subsets of cardinalities < 3 were deleted (e.g., the set <C,E> would be deleted if it preceded <C,E,G>), as were repeats. These processes (and those following) were implemented in the Python language using the music21 software package (as described in Cuthbert & Ariza, 2011).

The reduction process' input and parameters were defined as follows. The YCAC's lists of scale-degree sets were used both to provide the 2-gram probabilities within the reduction process and as the material for the reduction process itself. Any repeated chords that arose during the reduction process were conflated into a single chord event. The distribution of chords in the reduced series contained a long tail: the top 22 chords accounted for 85.3% of all chord occurrences, with the remaining chords accounting for only 14.7%. These infrequent chords were removed from the reduced series, and chord transitions were revised. This series was used to create a reduced vocabulary of scale-degree sets and a reduced 2-gram Markov model. The resulting vocabulary and syntax and its overlap with textbook chordal vocabularies are described in more detail in the online supplement.

The MIDI files of each excerpt in the Kostka-Payne corpus (Temperley, 2009) were chosen as the musical material to be analyzed. The corpus contains 46 examples of Western European common-practice music drawn from the Kostka-Payne textbook, ranging from the Baroque to Romantic era. Five pieces were removed due to difficulties converting them into digital representations usable by this study's methods. The ground truth was taken as the opening key indicated in the textbook instructor's edition.

To compare this model's behavior to other models, three key-profile implementations were used to assess the tonal orientations of the same examples: the Krumhansl-Schmuckler (Krumhansl 1990), the Temperley-Kostka-Payne (Temperley 2007), and Bellman-Budge (Bellman 2005) weightings. All three were available in music21's library of key-finding functions. Various window-lengths were attempted for these analyses, and it was found that using the first six offsets of each piece produced the best results.

2) Results. As shown in Figure 2, of the 41 pieces analyzed, the SPOKE's Reduced-YCAC model assigns the same tonic triad as the textbook 87.8% of the time. Of the 5 that did not overlap, twice the model judged the key to be the passage's relative major, twice the keys differed by fifth, and once the passage was too scalar for the model to recognize the correct underlying chords. Figure 2 also shows how often each key-profile model produced correct answers. The profiles of Krumhansl-Schmuckler, Temperley-Kostka-Payne, and Bellman-Budge produced rates of 78.0%, 85.4%, and 73.2% correct, respectively. (These findings track but differ from Albrecht and Shanahan (2013), likely because their specific implementation and music21's differ.)

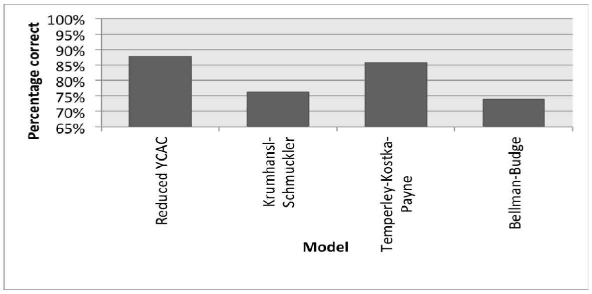


Figure 2: The percentage of key assessments on the Kostka-Payne corpus that agreed with the instructor's edition for each model

While this test addresses the output of key finding models, it does not test whether SPOKE executes the key-finding task in a way similar to a human, namely whether the model identifies chords, prolongations, and points of modulation in the same way a human would. The following test therefore compares the output of the analytical model to analyses of the same excerpts produced by undergraduate music students.

D. Test 2 – Comparing the Model to Human Annotations

1) Materials and Method. Of the Kostka-Payne excerpts whose key was successfully analysed by the reduced YCAC model, 32 were chosen to be analysed in depth by both humans and the model. The examples were selected to control for the passages' different characteristics, dividing into four categories with 8 examples in each group. Characteristics were labelled "Simple," "Modulating," "Chromatic," and "Chromatic Modulating." (These were determined by the author in relation to their placement in the Kostka-Payne textbook and the type of analytical knowledge they seemed designed to teach.)

The model's parameters were set as in Test 1, with several additions to allow for the entire excerpts to be analysed. For the automated analysis, the algorithm was run using a moving window designed to provide the model with a consistent number of non-repeating chords. The window began with four chords, but if the reduction process reduced that span to fewer than four chords, the window was extended until the process produced a four-chord analysis. The window then moved forward, progressing through the piece. Results were transcribed using a "voting" process. If three or more windows assigned on a particular scale-degree set at a given timepoint, the Roman-numeral annotation of that set was placed in the score at the appropriate timepoint. Using this voting process allowed the model to interpret modulations: for instance, if three earlier timepoints read a C-major triad as I, and three later timepoints analyse it as IV, then both those Roman numerals were placed under that moment in the score. If there was no agreement, a question mark was placed in the score. Since the final windows in the piece would only produce one or two annotations, the automated annotations ended before the example's final measure in several instances. If a series of chords was reduced to a single annotation, a line was used to indicate the prolongation of that single chord. (The reduced vocabulary included two non-triadic sets, the 11th and 21st most frequent chords in the vocabulary; since these chords were outliers and annotating a chord with a non-triadic structure would "give away" the origin of the analysis and unduly bias graders, these non-triadic structures were removed from the model's vocabulary.)

The 32 examples were also analysed by 32 undergraduates in their fourth and final semester of a music-theory sequence at the University of North Carolina at Greensboro's School of Music, Theatre and Dance. (Six subjects were drawn from the author's theory section, 26 were not.) Students were given 10 minutes to analyse their example, and were asked to use Roman numerals without noting inversion. The full instruction page, as well as all examples and analyses, can be found in the supplementary material.

Eight music theory faculty (each from different institutions, all currently teaching a music theory or fundamentals class) were then asked to grade a randomly selected group of 8 analyses, not knowing which of the set was algorithmically or human generated. The theorists were asked to provide a grade from 0 to 10 for each example, indicating the level of expertise in Roman-numeral analysis the annotations seem to convey (0 = a student with no music theory experience, 10 = the expertise of a professional).

2) Results. As shown in Figure 3, analyzing the average grades of the humans and the computer using a two-sided *t*-test is nearly significant (p=.07), indicating that the computer was given consistently higher grades but the variation renders the grades statistically indistinguishable. An Analysis of Variance (ANOVA) found no significant three-way interaction between the primary factors in this model (the grader, the example's characteristic, and whether the analysis had been produced by a human or computer) and the grades given to the examples ($F(1, 25) = .123$, $MSE = .615$, $p = .729$), nor was there an interaction between the grade and the computer/human producer ($F(1, 25) = .265$, $MSE = 1.324$, p=.611). The one significant effect was the interaction between grade and the example's characteristic ($F(3, 25) = 3.053$, $MSE = 15.27$, $p = .047$), due to the fact that "simple" excerpts received significantly higher grades.

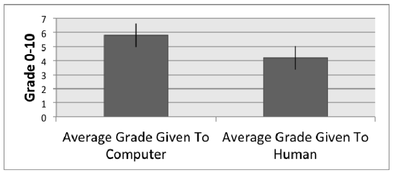


Figure 3. Comparing graders' assessments for both groups of analyses with standard error

E. Test 3 – Human or Computer?

1) Introduction, Materials, and Methods. While the previous tests show SPOKE to be performing well, it might behave in a way noticeably different from human behaviour. To investigate such potential differences, a final test was undertaken in which expert theorists were asked to distinguish between human and algorithmic analyses.

Both the human and automated analyses of Test 2 were used, now reordered into pairs. Each pair included the same excerpt analyzed twice, once by a human and once by the algorithm, with sequential and pairwise ordering randomized. Eight music theory faculty (each from different institutions, all currently teaching a music theory or fundamentals class, with two having participated in the earlier grading task) were presented with eight pairs of analyses. Each excerpt was analyzed twice by different graders. The task was to report which of the two they believed to be created by the computer and to provide a short written explanation of their choice. If the sorts of analytical choices made by humans were different than the algorithm's behaviors, the theorists would perform better than chance in their choices. (The full packet of analysis pairs appears in the online supplement.)

2) Results. Figure 4 shows the number of times the theorists correctly identified the computer-generated analysis, first shown as an overall percentage (in white) and then grouped by the excerpt's characteristic (in grey). Overall, the analyses were statistically indistinguishable: of the 64 choices made by the graders, only 35 were correct, falling well within the error of a $P(.5)$ binomial distribution. However, excerpts from the "chromatic" category were significantly distinguishable: of the 16 choices made by theorists, 75% (12) were correct ($p = .0384$).

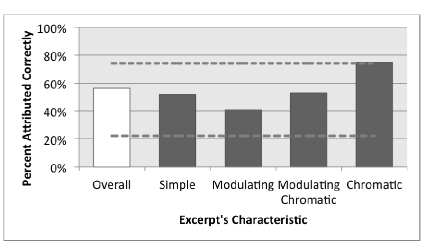


Figure 4. Dotted line shows the window for a $p < 0.05$ according to a $P(.05)$ binomial distribution

IV. DISCUSSION

These tests indicate that SPOKE, a chordal reduction process combined with an established analytical algorithm, seems to conform to human Roman-numeral annotations. Test 1 found the model to outperform key profiles when judging the opening key of excerpts within the Kostka-Payne corpus, while Tests 2 and 3 showed that analyses generated by the model were judged comparable to and – with the exception of analyses of chromatic passages – indistinguishable from those produced by undergraduate music majors. These findings suggest that a chordal key-finding process can, in fact, both draw its parameters from a musical corpus and also simulate the key finding behaviour and Roman-numeral annotations of humans.

The goal of this research is not to supplant pitch-based key finding models – their connections to music cognition are unassailable. However, these results suggest that tonal cognition might be a multifaceted and use multiple musical domains. For instance, a chord-based model loses its power in scalar passages and in monophonic music, while a chordal model might better perform in more thickly-textured music. Furthermore, the research shows that salient chordal information *can be learned by exposure to a corpus*. These results provide a fundamental proof-of-concept: if this model indeed captures the properties of common-practice music used to construct traditional triadic Roman-numeral vocabularies, then listeners may learn tonal vocabularies and analyse music in a way similar to how the model learns its vocabulary and assigns keys.

Nevertheless, these observations are preliminary and speculative. This model of tonal cognition seems to adhere to many behavioral aspects of the key-finding task, but the cognitive validity of much of its engineering remains to be more extensively tested against human behavior. After all, the value of computational work to music psychology is not only to model what is already known about some cognitive process, but to model what *might be true* given what we know about that cognitive process. Such empirically-based speculative models suggest new hypotheses, plausible explanations, and directions for future experimental research.

ACKNOWLEDGMENT

Many thanks to the expert graders: Damian Blattler, Rebecca Leydon, Patrick McCreless, David Crean, Bryn Hughes, Ben Crouch, Chris Brody, Megan Long, Drew Nobile, Andrew Aziz, Michael Buchler, Brent Auerbach, and Guy Capuzzo. I am grateful to David Sears, David Temperley, Trevor deClercq, and Josh Albrecht for comments on earlier drafts of this essay.

REFERENCES

Albrecht, J. & Shanahan, D. (2013). "The use of large corpora to train a new type of key-finding algorithm: An improved treatment of the minor mode." *Music Perception. 31*(1), 59-67.

Aldwell, E. & Schachter, C. (2011). *Harmony and voice leading* (4th ed.). Boston: Schirmer.

Barthelemy, J., & Bonardi, A. (2001). "Figured bass and tonality recognition." *Proceedings of the Second International Conference on Music Information Retrieval,* 129-135.

Bellman, H. (2005). About the determination of the key of a musical excerpt. *Proceedings of Computer Music Modeling and Retrieval*, 187-203.

Bharucha, J. J. (1991). Pitch, harmony and neural nets: A psychological perspective. In P. M. Todd & D. G. Loy (Eds.), *Music and Connectionism.* Cambridge, MA: MIT Press, 84-99.

Bharucha, J. J. (1987). "Music cognition and perceptual facilitation: a connectionist frame-work." *Music Perception,* 5/1, 1–30.

Brown, H.., & Butler, D. (1981). "Diatonic trichords as minimal tonal cue cells," *In Theory Only, 5*(6-7), 37-55.

Brown, H., Butler, D., & Jones, M. R. (1994). Musical and temporal influences on key discovery. *Music Perception, 11/4,* 371-407.

Browne, R. (1981). "Tonal implications of the diatonic set." *In Theory Only, 5*(6-7), 3-21.

Castellano, M. A., Bharucha, J. J., & Krumhansl, C. L. (1984). "Tonal hierarchies in the music of North India." *Journal of Experimental Psychology*: *General*, *113/3*, 394-412.

Creel, S. C., & Newport, E. L. (2002). "Tonal profiles of artificial scales: Implications for music learning." In C. Stevens, D. Burnham, G. McPherson, E. Schubert, & J. Renwick (Eds.), *Proceedings of the 7th International Conference on Music Perception and Cognition, Sydney,* 281-284.

Cuthbert, M, & Ariza, C. (2011). "music21: A toolkit for computer-aided musicology and symbolic music data," *Proceedings of the International Symposium on Music Information Retrieval*, 637-42.

Forte, A. (1973). *The structure of atonal music.* New Haven: Yale University Press.

Hörnel, D., & Ragg, T. (1996). "Learning musical structure and style by recognition, prediction and evolution." *Proceedings of the 1996 International Computer Music Conference*. San Francisco: International Computer Music Association, 59-62.

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. Cambridge, MA: The MIT Press.

Kostka, S., and Payne, D. (2012). *Tonal harmony with an introduction to twentieth-century music* (4th Ed.). New York: McGraw-Hill.

Krumhansl, C.L. & Shepard, R. N. (1979). "Quantification of the hierarchy of tonal functions within a diatonic context." *Journal of Experimental Psychology: Human Perception and Performance*, *S/4*, 579-594.

Krumhansl, C.L. (1990). *The cognitive foundations of musical pitch.* Oxford: Oxford University Press.

Laitz, S. (2008). *The complete musician: An integrated approach to tonal theory, analysis, and listening* (3rd Ed.). Oxford: Oxford University Press.

Matsunaga, R., & Abe, J. (2005). Cues for key perception of a melody: Pitch set alone? *Music Perception*, 23/2, 153-164.

Matsunaga, R., & Abe, J. (2012). "Dynamic cues in key perception." *International Journal of Psychological Studies*, 4/1, 3-21.

Pardo, B, & Birmingham, W.P. (1999). "Automated partitioning of tonal music," *Technical report, Electrical Engineering and Computer Science Department.* Ann Arbor: University of Michigan.

Quinn, I. (2010). "What's 'key for key': A theoretically naive key-finding model for Bach chorales." *Zeitschrift der Gesellschaft für Musiktheorie*, *7*(ii)*,* 151-163.

Roig-Francoli, M. (2010). *Harmony in context.* New York: McGraw-Hill Education.

Temperley, D. (2007). *Music and probability*. Cambridge: The MIT Press.

Temperley, D. & Marvin, E.W. (2008). "Pitch-class distribution and the identification of key." *Music Perception, 25*(3)*,* 193-212.

Temperley, D. (2009). "A statistical analysis of tonal harmony." http://www.theory.esm.rochester.edu/temperley/kp-stats/ (accessed 24 August 2015).

Tymoczko, D. (2011). *A geometry Of music: Harmony and counterpoint in the extended common practice*. Oxford: Oxford University Press.

White, C. (2013). "An alphabet-reduction algorithm for chordal *n*-grams." In *Proceedings of the 4th International Conference on Mathematics and Computation in Music*. Heidelberg, Springer, 201-212.

White, C. (2015). "A corpus-sensitive algorithm for automated tonal analysis." In T. Collins, D. Meredith, A. Volk (Eds.), *Mathematics and computation in music, LNAI, 9110* Springer, Heidelberg, 115-121.

A Corpus Approach to the Classification of Non-Chord Tones Across Genres

Claire Arthur[*1]

[*]*Music Dept., Ohio State University, U.S.A.*
[1]arthur.193@osu.edu

ABSTRACT

Music theorists distinguish between melodic notes that fulfill structural roles from notes that are considered elaborations or embellishments. Structural notes generally outline or form a part of the local harmony, while embellishing notes commonly are not members of the immediate harmonic structure. Thus, *non-chord tones* (NCTs) are a type of embellishment that fit outside of the local harmonic grammar. One method for identifying NCTs is to first identify harmonic changes, then tag melody notes as either members of the underlying harmony (chord tones) or as NCTs. However, within the classical tradition, there are compositional and stylistic norms for the use of melodic embellishing tones. As such, melodic features such as metric position, duration, and interval of approach and departure may function as identifiers of NCTs. How much information about NCTs is available from these melodic features? In popular music the relationship between melody and harmony is much freer, a phenomenon dubbed *the melodic-harmonic divorce*. How, then, might NCTs in popular music compare with usage in classical music?

Multiple corpora are used to investigate the usage and predictability of NCTs from purely melodic elements: A classical music corpus, consisting of Bach chorales, theme and variations by Mozart and Beethoven, and string quartets by Haydn; and a popular music corpus consisting of the McGill 'Billboard' corpus with melodic information added. The corpora are examined to compare usage of NCTs in classical versus popular music, as well as usage within each genre. Melodies are examined in isolation, and various algorithms attempt to classify melody notes as NCTs based on metric position, duration, and interval of approach and departure. To compare the analyses provided by the algorithms against some "standard," harmonic analyses are used as an alternative method for identifying NCTs. Implications for musical expectations and Music Information Retrieval are discussed.

I. INTRODUCTION

Which melodic factors might best predict whether a melodic tone is structural or embellishing? Do these factors change depending on the style and genre of music? When we hear a monophonic melody, for instance, are there certain features which help a listener understand the "correct" harmonic implications? By examining musical corpora which contain harmonic information, we can compare the predictive power of various melodic features – such as metric position, duration, and interval of approach and departure – in identifying non-chord tones (NCTs). That is, we can first examine the melody in isolation, using these features (both separately and in combination) to attempt to predict the location of NCTs. These predictions can then be compared against an alternative method of NCT identification: that of flagging any melodic tone which is not a member of the immediate underlying harmonic accompaniment.

II. BACKGROUND

The melody writing traditions in Western art music from the common practice period (henceforth simply referred to as "classical" music) and popular music, are, of course, quite different. Specifically, in classical music, there are "rules" for compositional practice dating back centuries. For example, most music theory texts classify various types of NCTs based on metric position and interval of approach and departure, with most NCTs resolving downward by step motion (Temperley, 2007). While some of these stylistic traits also appear in popular music, (e.g., suspensions, passing tones, etc.) the usage of NCTs in popular melodies appears much freer in comparison with the strict preparation and resolution of NCTs in classical melodies (Everett, 2001). Temperley (2007) specifically draws attention to the tendency for pop/rock melodies to exhibit a phenomenon dubbed "melodic-harmonic divorce" (attributed to Moore, 1995) where the melodic tones seem to be less governed by the local underlying harmony, and more likely to reflect the structure of the tonic (typically pentatonic) scale. In addition, Temperley noted that pop/rock melodies are less likely than classical melodies to have NCTs resolve down by step, are more likely to have multiple NCTs in a row, and in many cases contain NCTs that are left unresolved. This phenomenon is claimed to be more prominent in verse melodies than chorus melodies, with chorus melodies being more regulated by the harmonic structure (proposed as the "loose-verse, tight-chorus" (LVTC) model in Temperley, 2007).

Of course, the entire topic of NCTs brings to light the problem of defining what is meant by a "chord" (as discussed by Doll, 2013), and how chord events are determined in the corpora. While one might argue that any notes that sound concurrently can be considered a "chord," in classical music we typically think of chords as a triadic structure, or as a bass note with a combination of (mainly consonant) intervals above it. In popular music this issue is complicated further in that while triadic harmonies still feature prominently, chords with added 6ths, "sus" chords (a root, fifth, and suspended tone – typically a 2^{nd} or 4^{th} above the bass), and "power" chords (chords made up of only 5ths) occur frequently, and it calls into question the structural nature of a chord and what it means to consider a tone as "structural" or as "elaboration." Moreover, melodies and their harmonic accompaniments are typically not independent. As such, when analyzing a piece of music to determine the harmony (as was done in all of the corpora used in this study), typically the melodic notes are taken into consideration when determining the underlying harmony. Thus, while investigating the stylistic differences in melody between and across genres, this study will simultaneously help shed light on some of these contentious issues.

III. AIMS

Multiple corpora are examined to compare stylistic usage of NCTs in classical versus popular music. Specifically, a corpus of classical music consisting of Bach chorales, Haydn quartets, and theme and variations by Mozart and Beethoven, and a corpus of popular music consisting of the McGill "Billboard" corpus with melodic information added are compared and contrasted in terms of their usage of NCTs.

It is hypothesized that NCTs in classical music will be largely predicted by melodic factors, whereas NCTs (based on the traditional definition) in popular music will not be so easily identified. Within the genre of classical music it is hypothesized that theme and variations will have the greatest incidences of NCTs compared with quartets and especially chorales, due to the nature of variations as quintessential of embellishment.

Exploratory analysis will investigate which melodic factors carry the greatest predictive value for identifying NCTs both within and across genres. In addition, based on Temperley's (2007) model, it is hypothesized that the proportion of NCTs will be greater in verse sections than chorus sections in popular melodies. Finally, the question of whether popular melodies exhibit the tendency to move freely among the notes of the pentatonic scale (as Temperley, 2007 suggests) or whether the melodies tend to outline the tonic triad (as Nobile, 2015 suggests) can be directly evaluated.

IV. METHOD

From each of the corpora information from the melodies is extracted, such as the metric position of each note onset, the duration of each note, and the interval of approach and departure. All information is encoded in Humdrum format (Huron, 1995). An algorithm will be first applied to each melodic feature in isolation, and optimized to determine which feature in isolation is the best predictor of NCTs. For example, in classical music, based on rules of composition, a note that is lept to and from is likely to be a chord tone, regardless of metric position; however, ignoring melodic intervals, notes on strong beats are less likely to be NCTs than those on weak beats. Then, using the information from the combination of all melodic features, an algorithm will be optimized to predict the likelihood of a NCT in the melody. The success of the algorithm will be measured against harmonic analyses which were added manually to each corpus. Melodic tones will be classified as a member or non-member of the underlying harmonic accompaniment.

It is anticipated that algorithms for the most complex melodic corpus will work for less complex melodic corpora, but not vice-versa. In addition, it is predicted that the algorithm for identifying NCTs in the classical corpus will not work for the popular music corpus and vice-versa. That is, multiple algorithms are likely to be needed for the separate corpora.

As an example, the set of Bach chorales were compared against a test set of the theme and variations (specifically, just the themes) in their ability to predict a chord tone (CT) or NCT simply from the beat position of the given note (either 'on the beat' or 'off the beat'). As can be seen from Table 1, because of the frequent chord changes and relatively scarce off-beat melody notes in the chorales, this simple algorithm can correctly predict whether or not a note is a CT or NCT 83% of the time. However, when applied to the set of themes using only the beat position as a predictor, only 59% of CTs and NCTs were successfully predicted. However, the intervals of approach and departure may provide some additional information. As shown in Figure 1, the proportion of CTs approached or left by leap is larger than the proportion of CTs approached or left by step. Also, it is more likely that a note that is 'on the beat' and approached by step will be a CT than a note that is 'off the beat' and approached by step. If we then combine this information about the intervals of approach and departure with the beat position information and apply it to the set of themes – as shown in Table 1c – the model has improved its prediction rate and can now successfully predict CTs and NCTs at an accuracy rate of 74%.

Finally, because of the tendency for popular music to exhibit melodies and harmonies that are relatively "independent", it may be necessary to redefine "non-chord tone" status in popular music such that notes may be classified as more or less "stable" regardless of their relationship to the underlying harmony, and instead based primarily on melodic factors such as note duration, interval of approach and departure, and/or relation to the tonic triad. Note that this last factor would be consistent with Nobile's (2015) hypothesis that "the notes of the tonic triad…can under certain circumstances act as stable tones even if they are dissonant with the foreground harmony."

Note that in terms of the interval of approach and departure, the primary scale type will need to be considered. That is, as noted by Temperley (2007), what may appear to be resolution by "skip" may in fact simply reflect the fact that popular melodies are often within a pentatonic, rather than diatonic collection, and therefore a "step" must include a minor third.

Table 1: Proportion of predicted versus actual chord tones and non-chord tones for the Bach chorales and the Mozart and Beethoven themes.

a) Chorales: Prediction based on beat position(on/off)

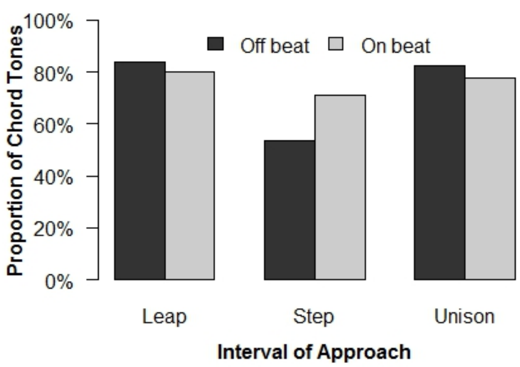
	NCT	CT
Predicted NCT	.07	.03
Predicted CT	.14	.76

b) Themes: Prediction based on beat position(on/off)

	NCT	CT
Predicted NCT	.12	.26
Predicted CT	.15	.47

c) Themes: Prediction based on beat position(on/off) + interval of approach & departure

	NCT	CT
Predicted NCT	.08	.07
Predicted CT	.19	.66



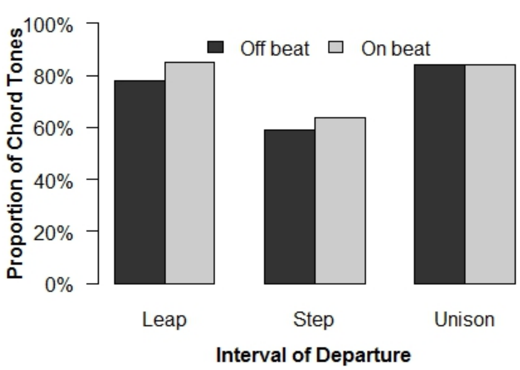


Figure 1. Proportion of chord tones and non-chord tones in the set of Mozart and Beethoven themes according to beat placement ('on the beat' or 'off the beat') and interval of approach and departure.

REFERENCES

Doll, C. (2013). Definitions of 'chord' in the teaching of tonal harmony. *Dutch Journal of Music Theory 18*(2), 91–106.

Everett, W. (2001). *The Beatles as Musicians: The Quarry Men Through Rubber Soul*. New York: Oxford University Press.

Everett, W. (2004). Making sense of rock's tonal systems. *MusicTheory Online 10*(4).

Huron, D. (1995). The Humdrum Toolkit: Reference manual.

Moore, A. 1995. The so-called ''flattened seventh'' in rock music. *Popular Music*, 14, 185–201.

Nobile, D. (2015). Counterpoint in rock music: Unpacking the "melodic-harmonic divorce." *Music Theory Spectrum (37),* 189-203.

Temperley, D. (2007). The melodic-harmonic 'divorce' in rock. *Popular Music 26*(2), 323-342.

Weisethaunet, H. (2001). Is there such a thing as the ''blue note''? *Popular Music* (20), 99–116.

Winkler, P. 1978. Toward a theory of popular harmony. *In Theory Only, 4*(2), 3–26.

Exploring the temporal capacity of memory for key-change music empirically and computationally

Érico Artioli Firmino[1,2], José Lino Oliveira Bueno[2], Carol Lynne Krumhansl[1]
[1]Department of Psychology, Cornell University, Ithaca, New York, United States of America
[2]Department of Psychology, University of São Paulo, Ribeirão Preto, São Paulo, Brazil

The literature contains different estimates of the temporal capacity of memory for music containing key-changes. We addressed this in three time reproduction experiments in which a modulating chord sequence was followed by instructions to produce a silent time interval with a computer stopwatch that matched the duration of the sequence just presented. Time estimates were affected by three variables: (1) the duration of each key, (2) the response delay (the time between the end of the sequence and the participant's response), and (3) the distance between the keys. The sequences in Exp 1 and 2 were 60 s in duration. Exp 1 sequences modulated from C to Gb and the relative duration of the two keys was varied. Exp 2 sequences modulated from C to Eb to Gb and the longest key was either the first, second, or third. Exps 1 and 2 showed that when the longest key was the first, the reproductions were shorter. However, when compared to non-modulating controls, short earlier keys still resulted in time underestimations. Exp 3 sequences were 20 s in duration, and contained modulations from C to F (close) or C to Gb (distant); the response delay was 25, 50, or 100 s. The farther the interkey distance, the shorter the time reproduction. All these results confirm predictions of the Expected Development Fraction (EDF) Model whereby time underestimates result from a mismatch between the actual and expected duration of key-changes (Firmino & Bueno, 2008). Exp 3 showed, in addition, that when the response delay was 25 s, the time reproductions were larger than the longer response delays. We further developed the EDF Model to take into account the greater interfering effect of extra-musical information (e.g., verbal, visual) on time estimates of music when the response delay is short. The model dynamically describes the long capacity of memory for key-change music throughout the stimulus-response period.

Form-bearing musical motives: Perceiving form in Boulez's *Anthèmes*

Cecilia Taher[1], Robert Hasegawa[1], and Stephen McAdams[1]
[1]Schulich School of Music, CIRMMT, McGill University, Montreal, Quebec, Canada

This paper combines experimental and theoretical methods to examine listeners' perception of post-tonal musical forms. Both the perceptual and theoretical literature highlight the importance of formal hierarchy. Nevertheless, whereas perceptual research has shown that listeners are sensitive to many hierarchical levels of post-tonal forms and that surface features are essential to form perception (e.g., Deliège, 1989), post-tonal theories have tended to disregard large-scale hierarchical levels and emphasize local-level abstract structures (e.g., pitch-class sets) instead of surface events. This marks a disconnect between perception and theory research agendas. Boulez intended to reconcile perception with composition, describing the form of his *Anthèmes* in terms of constantly changing yet clearly recognizable motives and large-scale perceptual cues (Nattiez et al., 2014). *Anthèmes* exists in two versions—for violin and violin with electronics—that share the same motivic families. In *Anthèmes 2*, the form is enlarged and the electronics create important timbral transformations.

Theoretical analyses of the form of *Anthèmes 1* and its enlargement in *Anthèmes 2* build on the findings from two experiments that explore musicians' perception of variations of selected motives representing the motivic families of *Anthèmes.* Participants hear each motive, subsequently listening to one version of *Anthèmes* and pressing a button when they recognize a transformation of the motive. The findings suggest that predictability (repeating successions of motives or electronic effects consistently associated with a motive), formal blend (motivic continuity and lack of surface contrast), and elapsed time affect listeners' apprehension of the hierarchy. Musical events that are predictable or associated with low formal blend lead listeners' attention to the local/indivisible level. Unpredictable and formally blended events and the passing of musical time shift attention to higher/divisible levels.

Learning the Language of Affect: A New Model That Predicts Perceived Affect Across Beethoven's Piano Sonatas Using Music-Theoretic Parameters

Joshua D. Albrecht*[1]

*Music Department, The University of Mary Hardin-Baylor, Belton, TX 76513, USA
[1]jalbrecht@umhb.edu

ABSTRACT

This paper examines affective expression in Beethoven's piano sonatas from a structural perspective by building multiple regression models using music-theoretic parameters as predictors of perceived affective expression. Thirty 5-second excerpts from the Beethoven piano sonatas were selected by polling musicians for those moments that were the most emotionally expressive in the repertoire. These excerpts were divided into a training set and a test set, matched by compositional period and inner/outer movement. 68 participants were randomly divided into one of the two groups in which they listened to each excerpt in that group and rated each on 9 affective dimensions. Each excerpt was analyzed for 14 different musical parameters, used as predictors in a regression model for each affective dimension. The models from the training set were applied to the same musical parameters for excerpts of the test set as a way of predicting participant ratings of affect. These predictions were tested against actual participant ratings as a way to test the generalizability of the models. Of the nine affective dimension models, six produced predictions that were significantly positively correlated with mean participant ratings in the test set, but all nine models were significantly better at predicting ratings than chance. The regression models were able to account for between 9-71% (Mean = 34.9%) of the variance in participant response. However, all nine reverse models, generated from the test set and tested on ratings from the training set were significantly positively correlated with mean participant ratings, and all were likewise better at predicting ratings than chance.

I. INTRODUCTION

Music and emotion has been the subject of scholarly inquiry for centuries, but recent empirical efforts have advanced the field in important ways with new methods and new types of insights into old questions. Specifically, recent research has discovered links between affective induction and personality (Rentfrow & Gosling, 2003; Ladinig & Schellenberg, 2011) and between musical characteristics and physiological responses like *frisson* (Sloboda, 1991). Some aspects of the induction and perception of musical affect appear to be shared cross culturally (Balkwill & Thompson, 1999; Balkwill, Thompson, & Matsunaga, 2004), although this effect may be more robust for physiological arousal than for valence (Egermann, *et al.*, 2015).

One promising line of research involves discovering connections between musical surface features and perceived affect. Juslin and Lindström (2011) found that particular musical features, such as pitch, rhythm, and articulation, tended to correlate with the perception of specific affective dimensions. Similarly, Schubert (2004) found strong correlations between surface musical features and perceived affect. Other models of affective expression have been built using Debussy's *Préludes* (Imberty, 1979), film music excerpts (Eerola, Lartillot, & Toiviainen, 2009), and television title themes (Tagg, 2006). However, the extent to which each of these studies' models apply to other musical styles remains unclear.

Despite this recent progress, much work remains. The way that musical structures elicit affective responses is not completely understood. Although many of the correlations between musical surface features and affective response appear to be cross-cultural, many of these correlations may be culturally bound or may even defy interpersonal generalization. For example, although the two styles may share some aspects of affective expression, it is likely that musical gestures conveying sadness in Baroque opera draw on significantly different types of musical structures than gestures of sadness in 1990's-era grunge rock.

In short, there may be different modes of expressing the same affective states in music between different genres, eras, or styles. Even within the musical dialect of one style or one composer, there may be different ways to express the same emotion. Therefore, it can be dangerous to overly generalize findings from the study of one particular set of excerpts to others. Although statistical methods can estimate the amount of overfitting in a regression model, it can still be difficult to know precisely how generalizable the results truly are. One way to address the question of how generalizable some experimental finding might be is to test the correlations found between musical gestures and perceived affect on a different dataset with a different set of participants.

II. AIMS

The purpose of this study is to develop a robust model correlating musical structures with perceived affective expression in a limited repertoire – the piano sonatas of Beethoven. Beyond simply correlating musical structures and perceived affect within one dataset, this study will test these generated models with a new dataset using different excerpts from the same repertoire. With this approach, the amount of overfitting present in the initial models can be directly examined, providing an estimate of the generalizability of the models to this specific repertoire.

III. METHODOLOGY

In prior work, Albrecht (2012) found that a model constructed from correlating perceived affect and musical surface features in the second movement of Beethoven's *Pathétique* sonata was able to predict perceived affect in other Beethoven sonata excerpts with some degree of accuracy. However, that model suffered from some important limitations. Specifically, the model was constructed for the affective expression of only one movement, limiting the independence of the data and the generalizability of the input of the model. Additionally, the affective labels were found to be problematic (Albrecht, 2014). In this study, a new set of affective labels were used, experimentally determined to be appropriate for the Beethoven piano sonatas (see Albrecht, 2014), and the model was constructed from several different movements with a greater variability of affective expression.

A. Affective Categories

Zentner, *et al.* (2008) argues that the best paradigm for studying musical affect is to use an eclectic list of affect terms specifically suited to the music under investigation. Albrecht (2014) conducted a series of three studies with different methodologies to empirically derive an eclectic list of affective dimensions suitable for studying the piano sonatas of Beethoven. The results of that study suggest nine eclectic affect terms: angry, anticipating, anxious, calm, dreamy, exciting, happy, in love, and sad. While these terms cover many of the common affective dimensions (relating to both the basic emotions and the arousal/valence paradigm), several less common terms are also present, such as dreamy and in love. The terms were also chosen to be largely orthogonal. This paper uses these nine terms.

B. Stimuli

Of primary importance to this study is the selection of excerpts to investigate. In the selection process, there are three primary considerations: which excerpts to select, how long the excerpts should be, and which recordings should be used.

One sampling method would involve the principal investigator selecting excerpts deemed to exhibit strong affective expression. However, this approach may introduce researcher bias. A second approach of randomly sampling from the repertoire might suffer from selecting affectively neutral excerpts, such as Alberti bass passages. To avoid these problems, three sources of expert musicians were polled: The Society for Music Theory listserv, the American Musicological Society listserv, and the Piano Street online forum. In these posts, respondents were asked to submit specific measures or excerpts of "the most emotionally expressive moments in the Beethoven sonata literature." As a result of this query, 30 excerpts were attained, divided into a training set and a test set. The two sets were matched for both compositional period (early, middle, late) and inner/outer movement. The excerpts attained from the poll are shown in Table 1.

To try to minimize the effect of performer bias, recordings were selected randomly from the compact disc holdings of the University of Texas at Austin. In total, there were 155 Beethoven piano sonata discs or collections. Of these, historical recordings with poor quality were eliminated, as were any recordings by "Joyce Hatto." No more than one recording was selected per CD. The performers and year of release of chosen recordings are provided in Table 2.

A final consideration is how long the excerpts should be. Respondents' descriptions of their suggested excerpt are provided parenthetically in Table 1. Some of the passages described are very short and specific (specific measures) while others are very long (themes). To control for excerpt length, and in keeping with Albrecht (2012), excerpts of five seconds in length were chosen. If the excerpt provided by the respondent was longer than five seconds, the five seconds including the moment most closely matching the description was chosen. If there was no description of the passage, the first five seconds were used.

Table 2. Thirty recordings chosen randomly from 155 compact disc recordings at the University of Texas at Austin's library.

Performer	Date	Performer	Date
Badura-Skoda, Paul	1988	Landowski, Wanda	1927
Barenboim, Daniel	1984	Levin, Beth	2013
Benedetti, Arturo	?	Lortie, Louis	1994
Brendel II, Alfred	1998	Lortie, Louis	1998
Chodos, Gabriel	2007	Michelangeli	1960
Fischer, Edwin	1987	Nissman, Barbara	2007
Goode, Richard	1991	Perahia, Murray	1986
Haskil, Clara	1960	Ratusiñski, Andrzej	1981
Horszowsky, Mieczyslaw	1991	Richter, Sviatoslav	1960
Jandó, Jenö	1987	Richter, Svjatoslav	1971
Katahn, Enid	2000	Roberts, Bernard	1988
Kovacevich, Steven	1960	Rose, Jerome	1996
Kovacevich, Steven	1960	Rosen, Charles	1989
Kovacevich, Stephen	2001	Wild, Earl	1987
Kuerti, Anton	1975	Vogel, Edith	?

Table 1. Thirty excerpts attained from a survey of the Society for Music Theory listserv, the American Musicological Society listserv, and the Piano Street forum. Respondents were asked for the "most emotionally expressive excerpts" in the Beethoven piano literature, and to list the measures, themes, or sections along with the movements (descriptions are provided in parentheses). Movements were paired by compositional period and outer/inner movement, and then were randomly assigned to the training set or the test set.

Training Set			Test Set		
Early Period	Inner Movement	Op. 2, No. 1, II (2nd theme)	Early Period	Inner Movement	Op. 7, II (2nd theme)
		Pathétique, Op. 13, No. 1, II (entire mvt.)			Op. 10, No. 3, II (development)
	Outer	Op. 2, No. 1, IV (mm. 20-50)		Outer	Op. 10, No. 3, I (no section identified)
Middle Period	First Movement	Op. 26, I (resolutions in Var. III)	Middle Period	First Movement	Op. 28, I (mm. 109-127, climax of A chord)
		Op. 27, No. 2, I (transition to the A1 section)			Op. 31, No. 2, I (recit. at close of movement)
		Op. 31, No. 3, I (ms. 99)			Appassionata, Op. 57, I (start of 2nd theme)
		Waldstein, Op. 53, I (beginning of 2nd theme)		Last Movement	Moonlight, Op. 27, No. 2, III (opening theme)
	Last Movement	Op. 26, IV (main theme)			Waldstein, Op. 53, III (major 7th arpeggios)
		Tempest, Op. 31, No. 2, III (mm. 169-173)			Les Adieux, Op. 81a, III (ending, new tempo)
Late Period	First Movement	Op. 109, I (retransition)	Late Period	First Movement	Hammerklavier, Op. 106, I (beginning)
		Op. 110, I (sad part)			Op. 111, I (beginning)
	Middle	Op. 81a, II (2nd theme)		Middle Movement	Hammerklavier, Op. 106, II (beginning)
	Last Movement	Op. 90, II (canon before the conclusion)			Op. 101, III (intro to final movement)
		Hammerklavier, Op. 106, III (mm. 60 or 145)		Last Movement	Op. 109, III (Var. VI)
		Op. 111, II mm. (116-121)			Op. 110, III (mm. 21-23)

C. Participants

This study involved 68 participants from the University of Mary Hardin-Baylor, consisting of freshmen and sophomore music majors. These participants were randomly divided into two groups, a group for training regression models and a group for testing the regression models. Across both groups, there were 33 females, a mean age of 19.4 (SD = 1.5), a mean number of years of musical training of 7.1 (SD = 3.7), and a mean Ollen Musical Sophistication Index (Ollen, 2006) of 455.0 (SD = 243.9). As a rough gauge of the extent of the participants' familiarity with Beethoven's piano sonatas, participants were also asked how familiar they were with early Romantic piano music and how often they listened to it (from 1-5). The mean across both groups was 2.6 (SD = .95) and 2.8 (SD = 1.1), respectively. There were no significant differences between the two groups regarding gender, age, years of musical training, OMSI score, or familiarity with or frequency of listening.

D. Procedure

Subjects participated in groups of up to 8 in a computer lab. After reading the experiment's instructions, participants listened to two trial excerpts to ensure they understood the instructions and interface. Participants then listened to all fifteen excerpts from their set and rated each one along nine different affective dimensions, including one extra trial per affect to test for intra-subjective reliability. Participants listened to the excerpts on headphones adjusted to a comfortable listening level, and rated each excerpts affective expression on a 7-point Likert scale.

IV. RESULTS

A. Reliability

Each participant rated a randomly-chosen excerpt twice for each affective dimension resulting in 15 pairs of matched ratings. Seven participants scored lower than the *a priori* cut-off set to a correlation of +.7, and so were eliminated.

B. Affective Exemplars

Average affect ratings were calculated for each excerpt for each affective dimension. Table 3 shows the excerpts that had the highest average ratings for each of the affective dimensions. These excerpts might be construed as exemplars of each affective dimension among this set of thirty excerpts.

Table 3. The eliciting the highest rating for by affective scale.

Affective dimension	Highest rated excerpt	Mean rating (SD)
Angry	Moonlight, Op. 27, No. 2, III (mm. 1-4)	6.3 (.9)
Anticipating	Tempest, Op. 31, No. 2, III (mm. 169-173)	6.2 (1.3)
Anxious	Moonlight, Op. 27, No. 2, III (mm. 1-4)	6.1 (1.4)
Calm	Pathétique, Op. 13, No. 1, II (mm. 1-2)	6.5 (.7)
Dreamy	Waldstein, Op. 53, III (mm. 251-254)	6.5 (1.1)
Exciting	Op. 10, No. 3, I (mm. 133-139)	6.8 (.5)
Happy	Op. 31, No. 3, I (ms. 99)	6.5 (.8)
In love	Hammerklavier, Op. 106, III (ms. 60)	5.9 (1.2)
Sad	Op. 101, III (ms. 1)	6.5 (1.1)

C. Music-Theoretic Predictors

Each of the 15 excerpts in the training set were analyzed for fourteen musical parameters. Table 4 shows each musical parameter and the way that each was operationalized. The score and recording were both consulted for each 5" excerpt.

Table 4. Fourteen regression parameters used in the model and their operationalization.

Regression parameter	Operationalization
Dynamic level	Performed relative dynamic level: Coded *pp–ff* (1-5)
Crescendo	Crescendo or decrescendo: Coded "strong dim. – strong cresc." (-2–2)
Density	Maximum number of concurrently attacked
Surface rhythm	The total number of onset moments
Tempo	Performed relative tempo: Coded 1-5
Articulation	Performed articulation: Coded 1-5 (1 = legato)
Pitch height	Two separate values: Highest and Lowest pitch in semitones from middle C
Pitch direction	Correlation of pitch height/time: Coded -2-2 (all descending- all ascending)
Tendency tone	Presence of tendency tones: Coded 0/1/2 (0 = none; 1 = LT or Dom. 7, 2 = suspension or melodic chromatic half-step motion)
Mode	Predominantly major or minor harmonies: Coded -1-1 (1 = major; 0 = neither)
Harmonic dissonace	Presence of harmonic dissonance: Coded 0-2 (0 = only M or m triads; 1 = Dom. 7; 2 = dim. or other dissonant/chromatic harmonies)
Harmonic tempo	Number of harmonies per segment
Closure/ cadence	Presence of closure/cadence: Coded 0-2 (2 = authentic; 1 = other cadence)

D. Predictor Selection

The fourteen music-theoretic predictor variables were then correlated with participant responses for each of the nine affective categories using a multiple regression model. Model selection is not a straightforward task, and suffers from dangers of overfitting and covariance. Regression models were selected that a) maximized variance accounted for, b) minimized overfitting by limiting the number of predictors, and c) selecting for theoretically meaningful predictors. One model was created for each affective dimension. Models included 3-4 predictor variables each, accounting for between 21.1%-70.0% (mean of 46%) of the variance in participant ratings. The musical parameters used in each model, along with the directionality of the effect, are shown in Table 5.

Table 5. Regression models derived from the training set. Adj. R^2 values are what percent of the variance is explained.

Affective dimension	Regression model	Adj. R
Angry	Louder dynamic, minor mode, lighter texture	0.506
Anticipating	More staccato, lighter texture, minor mode	0.211
Anxious	Louder dynamic, minor mode, more staccato, lighter texture	0.391
Calm	Quieter dynamic, more legato, major mode	0.630
Dreamy	Slower harm tempo, legato, major, crescendo	0.493
Exciting	Faster rhythms, faster tempo, more staccato	0.700
Happy	Faster rhythms, descending contour, major	0.427
In love	Decrescendo, descending contour, major	0.348
Sad	Slower rhythms, legato, minor mode, slower harmonic tempo	0.439

E. Affect Predictions

To test the generalizability of these models, each of the fourteen musical parameters used to construct the nine training set models were measured for the fifteen excerpts in the test set. These parameters were fed into each affect's regression model, resulting in a prediction in 'Likert space' for each affective dimension for each excerpt in the test set. Importantly, both the excerpts used and the participants tested were new for the test set. As a result, the proximity of the model predictions to the participant ratings provides a metric for whether the musical structures associated with affective expression are generalizable across the Beethoven piano sonatas and across listeners. Fig. 1 shows a sample comparison between predictions and rated affects of Op. 109, I and Op. 81a, III in the test set. Red asterisks show predictions made by the training set models and dots surrounded by the blue 95% confidence intervals show the mean participant ratings of the excerpt for the given affective dimension.

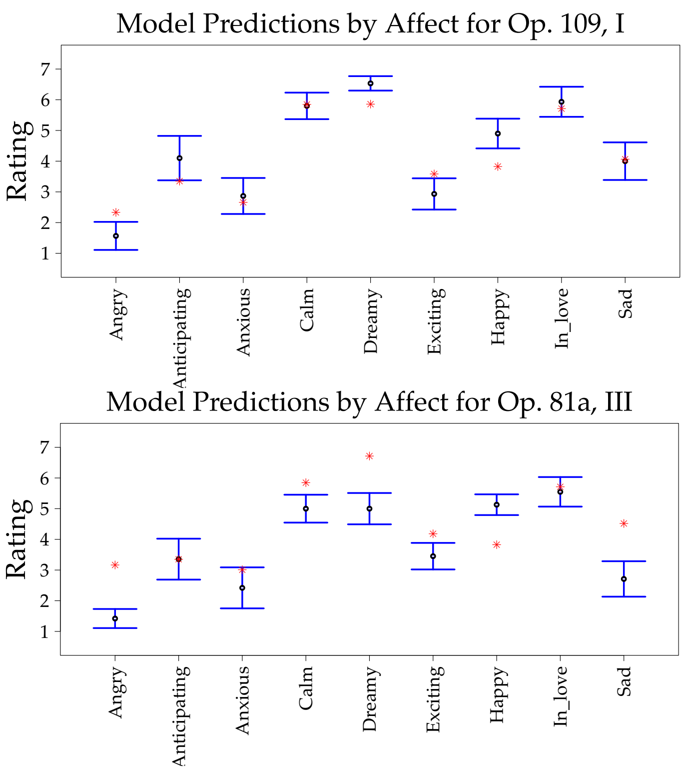


Figure 1. Model predictions for Op. 109, I and Op. 81a, III in the test set. Red asterisks show predictions from the training set regression models based on the musical parameters measured in these two test set excerpts plotted against the 95% confidence intervals around the mean of actual participants ratings from the test set.

The sort of 'affective profiles' provided in Fig. 1 are interesting in themselves and would be useful in analyzing the perceived affective expression of each excerpt. For the purpose of this study, however, it is more important to test the generalizability of the models derived from the training set by examining the relationship between those models' predictions (red asterisks) and the actual participant ratings of affective expression (confidence intervals). To illustrate, consider Fig. 2, in which the plot of model predictions and participant ratings for the exciting and sad affective dimensions are plotted.

Figure 2 provides a profile not of an excerpt but of an affective model. In short, the distance between the predictions and participant ratings provides a metric of model success and, consequently, generalizability of affective expression across the repertoire. Although most of the asterisks are relatively close to the participant ratings, some of the model predictions are fairly distant from attained ratings, especially for the sad affect model.

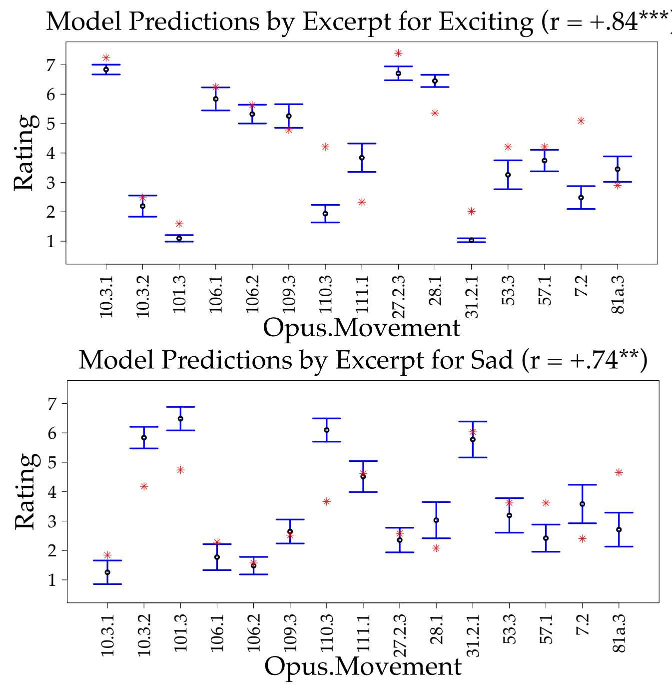


Figure 2. Model predictions for two of the nine affective dimensions: exciting and sad. As in Fig. 1, red asterisks represent model predictions plotted against 95% confidence intervals. Although all three models shown are significantly correlated with participant ratings, some excerpts reveal slight inaccuracies in prediction.

One conservative way of testing whether or not the predictions made by the models are significantly predictive of participant ratings is to correlate the models' predictions with the mean of participant ratings for each excerpt. In some ways, this test is greatly underpowered, because it only compares two sets of fifteen estimates. Nevertheless, under a correlation test using Pearson's *r*, six of the nine affective dimensions were significantly correlated at $p < .05$, including calm, dreamy, exciting, happy, in love, and sad. The correlations obtained and the amount of variance explained are shown in the center column of Table 6.

Not surprisingly, the amount of variance that the regression models are able to explain in the test set are generally lower than in the training set, ranging from only 9%-71% (averaging only 34.9%). While four of the models perform just as well or better on the test set than on the training set – exciting, happy, in love, and sad – the rest are able to account for less variance in participant rating. This is a predictable effect of model overfitting and suggests that some of the models do not generalize as well as others to the entire repertoire. Nevertheless, the fact that several of the models accurately predict participant responses is consistent with the idea that those affective dimensions can be reliably modeled on the Beethoven piano sonatas at large.

Table 6. Accuracy metrics of each training set regression model

Affective dimension	Correlations (R predictions and mean participant ratings	Mean distance (SD) closer to rating than 'noise distribution'
Angry	$r = +.71$ (.10), $p = .02$	1.15 (1.6), $p < .0001$
Anticipating	$r = +.50$ (.09), $p = .02$	0.65 (1.5), $p < .0001$
Anxious	$r = +.53$ (.16), $p = .02$	0.66 (1.5), $p < .0001$
Calm	$r = +.82$ (.40), $p = .0004$	1.39 (1.8), $p < .0001$
Dreamy	$r = +.70$ (.28), $p = .005$	1.30 (1.6), $p < .0001$
Exciting	$r = +.87$ (.71), $p = .0001$	1.73 (1.4), $p < .0001$
Happy	$r = +.81$ (.58), $p = .0002$	1.27 (1.6), $p < .0001$
In love	$r = +.48$ (.27), $p = .002$	0.69 (1.5), $p < .0001$
Sad	$r = +.77$ (.55), $p = .003$	1.16 (1.8), $p < .0001$

Another way of testing whether or not the regression models can accurately predict participants' response is suggested by the null hypothesis, which states that the relationship between musical structures and participant responses in one set of excerpts would be unrelated to the relationship in another set. If that is the case, then the predictions from one model applied to the predictors in a different set of excerpts would provide a poor fit equivalent to random noise rather than good predictions and should not be significantly different from participant responses generated by any other unrelated model.

This hypothesis can be directly tested by examining the distances in 'Likert space' between model predictions and participant responses. Specifically, random noise distributions can be generated by measuring the distance on the 7-point Likert scale between predictions of different affects' models and matching affect participant ratings. For example, the predictions made by every affect model *but* the angry model could be compared to participant responses for the angry affect. The Likert distances between the unmatched model predictions and participant ratings for each affect can be made into a 'noise distribution,' which can then be compared to a distribution of distances between the angry model predictions and angry participant ratings.

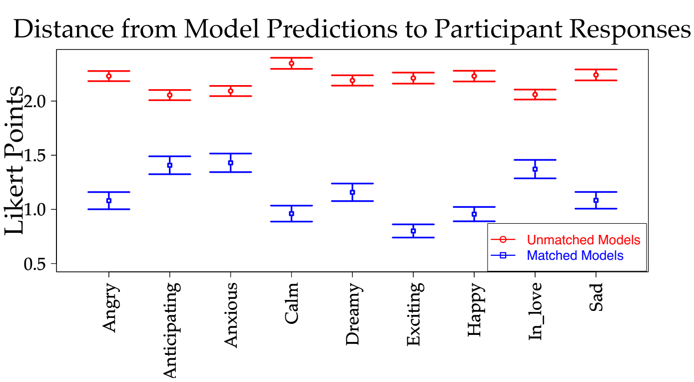


Figure 3. Distance, in Likert points on a 7-point scale, between model predictions and participant responses. On this scale, the smaller the distance, the more accurate was the prediction. The blue distances represent matched affect model/response comparisons, and the red distances represent the distances between model predictions for different affects than the participant's rated for each affect. All distances are significantly closer at *p* < .0001.

The results from this test are given in the right column of Table 6. Figure 3 displays the results graphically. Specifically, distances between participant ratings and matched model predictions for each affect are shown in blue and distances between ratings and unmatched model predictions are shown in red. In Fig. 3, the smaller the Likert distance, the closer the prediction and more successful the model. As can be seen in Table 6 and Fig. 3, matched model predictions were significantly more accurate at predicting participants' responses than chance, as measured by unmatched model prediction distances.

V. DISCUSSION

For each affective dimension, the regression models generated from one set of participants on the training set excerpts predicted ratings from new participants on new excerpts at rates better than chance. However, a further question could be raised about regression models that could be generated from the *test* set. In other words, given that the data from each set were collected independently, the study could theoretically be run backwards, in which the regression models generated from the test predict participant responses from the training set.

It should be mentioned here that this reversal is *post hoc* and after the models were built, tested, and examined. Caution should be advised in the treatment of the following data, in the first case because the author seeing the first set of models may bias the construction of the second set of models. In the second case, the way that the two sets of regression models interact is likely highly correlated; even though the predictors chosen may be different, if a model generated from dataset A performs well on dataset B, then it would not be surprising for the dataset B model to perform well on dataset A. Therefore, the following further examination of the data should be taken as *post hoc* and care should be taken in its interpretation.

A. Model comparison

Nine new regression models, one per affective dimension, were generated for the 'test set' dataset in a similar way as the 'training set' models were constructed (see section IV: D). These models, along with the amount of variance they account for, are shown in Table 7. The test set regression models used 2-4 predictors each, accounting for between 10.3%-67.1% (mean of 47.1%) of the variance in participant ratings.

Again, variable selection is not an exact science. The variables included in the model were chosen to be as orthogonal as possible, to explain as much of the variance in participant ratings as possible, and to do so with as few variables as possible. In many cases, the variables included as predictors mirrored the original set of models, but more predictors were unique than shared. Of the variables shared in common between both training and test models for an affect, those that agree in directionality are shown in **bold**. Those that disagree in directionality are shown in *italics*. Finally, those predictors that are analogous (typically surface speed and tempo) are shown by underline. As can be seen by comparing Tables 5 and 7, many of the predictor variables chosen in these new models agree in direction with the training models.

Table 7. Regression models derived from the test set. Adj. R^2 values are what percent of the variance is explained.

Affective dimension	Regression model	Adj. R
Angry	More tendency tones, **minor mode**, higher highest pitch	0.378
Anticipating	More dissonance, **minor mode**, higher highest pitch	0.203
Anxious	Higher highest pitch, faster rhythms, more tendency tones, *more legato articulations*	0.229
Calm	**Quieter dynamic**, slower tempo, higher lowest pitch	0.602
Dreamy	Slower tempo, lighter texture, fewer tendency tones, *faster harmonic tempo*	0.401
Exciting	Louder dynamic, **faster tempo**	0.671
Happy	Faster tempo, more staccato, **major mode**	0.589
In love	*Crescendo*, **major mode**, less dissonance	0.103
Sad	Quieter dynamic, more tendency tones, slower tempo, minor mode	0.579

Comparison of the two sets of models reveals patterns in predictor variables that are generally unsurprising. However, the fact that only 6 of 29 variables used in all nine test set models are the same and influence results in the same direction underscores the great variance involved in a study like this. In the first case, these results may reveal different modes of expressing the same affective state or the difficulty of generalizing across many excerpts. Moreover, it might be the case that the group of 34 participants represented in the test set may have been attending to different musical elements than the 34 participants in the training set to make their determinations. Again, the relatively accurate predictions made by the models are remarkable given the sizable differences between the predictors chosen by the different models, and is consistent with the notion that composers use quite a bit of redundancy in the communication of affective signals.

B. Reverse Model Predictions

To complete the reverse comparison, predictions were made by the 'test set' models on the 'training set' data and compared against the participant ratings. These results are shown in Table 8.

Table 8. Accuracy metrics of each test set regression model

Affective dimension	Correlations (R predictions and mean participant ratings	Mean distance (SD) closer to rating than 'noise distribution'
Angry	$r = +.59$ (.35), $p = .02$	1.15 (1.6), $p < .0001$
Anticipating	$r = +.58$ (.34), $p = .02$	0.65 (1.5), $p < .0001$
Anxious	$r = +.58$ (.34), $p = .02$	0.66 (1.5), $p < .0001$
Calm	$r = +.79$ (.62), $p = .0004$	1.39 (1.8), $p < .0001$
Dreamy	$r = +.67$ (.45), $p = .006$	1.30 (1.6), $p < .0001$
Exciting	$r = +.84$ (.71), $p = .0001$	1.73 (1.4), $p < .0001$
Happy	$r = +.82$ (.67), $p = .0002$	1.27 (1.6), $p < .0001$
In love	$r = +.74$ (.55), $p = .002$	0.69 (1.5), $p < .0001$
Sad	$r = +.70$ (.49), $p = .004$	1.16 (1.8), $p < .0001$

The middle column of Table 8 shows correlations between the means of each set of participant ratings and the model predictions. Again, these correlations only involve two sets of 15. Nevertheless, each set of predictions is significantly positively correlated with participant responses. These test set model predictions are better correlated with the training set ratings than the reverse (see IV: E). One possible reason for this result might be because the participants in the test set provided less 'noisy' responses. Another potential reason could be that the excerpts in the test set covered a wider range of affective expression and so the models derived from the data are more generalizable. Yet another possible explanation could be that the model selection process was biased by information from the first dataset.

As another test, Likert distances were calculated between affect ratings and matched model predictions and between the 'noise distribution' of unmatched model predictions. These results are shown in the rightmost column of Table 8 and graphically in Figure 4. As with the training set, Likert distances for regression models were significantly closer than chance.

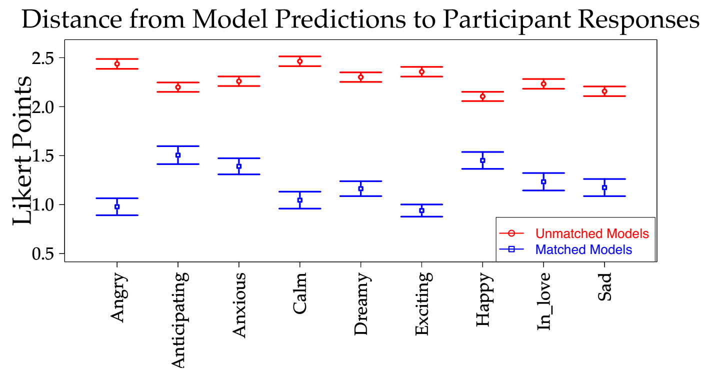


Figure 4. Distance, in Likert points on a 7-point scale, between model predictions and participant responses. This figure is the conceptual reverse of Figure 3. Specifically, the distance between test set model predictions was measured against participant ratings in the training set. Again, the blue distances represent matched affect model/response comparisons, and the red distances represent the distances between model predictions for different affects than the participant's rated for each affect. All distances are significantly closer at *p* < .0001.

C. Supermodel construction

As with the training set models, the test set models were significantly predictive of participant ratings in the training set. These results are consistent with the hypothesis that the regression models do not involve excessive overfitting, but instead that they truly capture some of the musical structures correlated with affective expression in the Beethoven piano sonatas. The high levels of accuracy in predictions are particularly noteworthy in light of how different the models are between the training and test sets in terms of the predictor variables used.

As a result of the inter-model accuracy of the models' predictions, it is appropriate to consider the entire set of 30 excerpts as one large dataset. From this large dataset, one supermodel for each affective dimension can be created that will explain the musical structures associated with affective expression in ways that are even more generalizable.

One benefit of using the larger dataset is that there is more variance in the way that each affective dimension is represented. Coupled with the increased size of the dataset, the extra variance allows for the statistical inclusion of more predictor variables that are significantly correlated with affective expression. As a result, the supermodels promise to more fully model the way that affective expression in this repertoire is communicated in different ways and a deeper understanding of nuances in the relationship between musical structures and that affective expression.

The supermodels are shown in Table 9. As in Tables 5 and 7, the musical parameters used in each supermodel along with the effect of the directionality are shown in the center column of Table 9. The adjusted R^2 values are in the right column, showing how much of the variance in participants' ratings the models are able to explain. Of note is the tendency for adjusted R^2 values to decrease from the smaller data set models. This effect is largely due to the increase of variance in the total dataset. Because the bigger dataset contains excerpts with a greater range of expression for each affective dimension and because there is a greater number of human raters, less of the variance can be accounted for. However, as a direct consequence, it is likely that the supermodels would be better at predicting new listeners' ratings for new excerpts from the Beethoven piano sonatas. Knowing the extent to which these models can successfully predict new listeners' ratings for new excerpts must await direct testing in a new experiment.

Table 9. Regression supermodels derived from the entire data set. Adj. R^2 values are what percent of the variance is explained.

Affective dimension	Regression model	Adj. R
Angry	**Minor mode**, **more tendency tones,** lower lowest pitch, **louder dynamic**	0.353
Anticipating	**More dissonance, minor mode**, **more staccato, higher highest pitch**	0.168
Anxious	More dissonance, **minor mode,** more staccato, **higher highest pitch**	0.238
Calm	**Quieter dynamic**, **slower tempo**, less dense texture, lower highest pitch	0.539
Dreamy	**Slower tempo**, quieter dynamic, **fewer tendency tones**, more dissonance	0.402
Exciting	**Faster surface speed, faster tempo, louder dynamic, staccato articulations**	0.678
Happy	**Major mode, faster tempo, staccato articulations, descending melodies, fewer tendency tones**	0.493
In love	**Major mode,** fewer tendency tones, higher lowest pitch, crescendo	0.193
Sad	**Slower tempo, slower surface rhythms, more legato, minor mode**	0.457

VI. CONCLUSION

Although the predictions of the regression models generated from the training set did not always fall within the confidence intervals of the participant ratings of the training set, the results are consistent with the hypothesis that the models capture some generalizable elements of affective expression, at least in Beethoven's piano sonatas. Moreover, testing the reverse regression models provided even closer estimates of listener perception of affective expression. Despite differences in the model structures as seen in predictor variable selection, regression models were significantly better than chance in every case. Given the similarity between test set and training set models, the *post hoc* supermodels may provide the best approximation of a generalizable estimate of musical structures associated with affective expression in the Beethoven piano sonatas. Again, further testing on a new dataset is necessary before being able to ascertain how accurate these models might be.

ACKNOWLEDGMENT

I would like to thank Aaron Baggett for spending several of his valuable hours talking with me at length about the unique statistical problems involved in this study. He not only provided a great sounding board for thinking through these issues, but was an enthusiastic conversation partner, and for that I am deeply grateful.

REFERENCES

Albrecht, J. (2014). Naming the abstract: Building a repertoire-appropriate taxonomy of affective expression. In *Proceedings of ICMPC-APSCOM 2014 Joint Conference* (pp. 128-135). Seoul, South Korea: Yonsei University.

Albrecht, J. (2012). A model of perceived musical affect accurately predicts self-reported affect ratings. In *Proceedings of 12th International Conference of Music Perception and Cognition* (pp. 35-43). Thessaloniki, Greece: Aristotle University.

Balkwill, L. L, & Thompson, W.F. (1999). A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perception: An Interdisciplinary Journal, 17*, 43-64.

Balkwill, L. L., Thompson, W.F., & Matsunaga, R. (2004). Recognition of emotion in Japanese, North Indian, and Western music by Japanese listeners. *Japanese Journal of Psychological Research, 46*, 337-349.

Eerola, T., Lartillot, O., & Toiviainen, P. (2009). Prediction of multidimensional ratings in music from audio using multivariate regression models. In *Proceedings of 10th International Conference on Music Information Retrieval (ISMIR 2009)* (pp. 621-626). Kobe, Japan.

Egermann, H., Fernando, N., Chuen, L, & McAdams, S. (2015). Music induces universal emotion-related psychophysiological responses: Comparing Canadian listeners to Congolese Pygmies. *Frontiers in Psychology, 5*, 1341. doi: 10.3389/fpsyg.2014.01341

Imberty, M. (1979). *Entendre la musique*. Paris: Dunod.

Juslin, P., & Lindström, E. (2011). Musical expression of emotions: Modeling listeners' judgments of composed and performed features. Special Issue: Music and Emotion. *Music Analysis, 29*(1-3), 334-364.

Ladinig, O. & Schellenberg, G. (2011). Liking unfamiliar music: Effects of felt emotion and individual differences. *Psychology of Aesthetics, Creativity, and the Arts, 6*, 146-154.

Ollen, J. (2006). *A Criterion-related Validity Test of Selected Indicators of Musical Sophistication Using Expert Ratings* (Doctoral dissertation). Columbus, OH: Ohio State University.

Rentfrow, P. J. & Gosling, S. D. (2003). The do re mi's of everyday life: The structure and personality correlates of musical preference. *Journal of Personality and Social Psychology, 82*, 1236-1256.

Schubert, E. (2004). Modeling perceived emotion with continuous musical features. *Music Perception: An Interdisciplinary Journal, 21*(4), 561-585.

Sloboda, J. (1991.) Music structure and emotional response: Some empirical findings. *Psychology of Music, 19*, 110-120.

Tagg, P. (2006). Music, moving images, semiotics, and the democratic right to know. In Ulrik Volgsten (Ed.), *Music and Manipulation: On the Social Uses and Social Control of Music* (pp. 163-186). New York, NY: Berghahn Books.

Zentner, M. R., Grandjean, D., & Scherer, K. R. (2008). Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion, 8*, 494-521.

Modelling melodic discrimination using computational models of melodic similarity and complexity

Peter M.C. Harrison[1], Daniel Müllensiefen[2]
[1]School of Electronic Engineering and Computer Science, Queen Mary, University of London, UK
[2]Department of Psychology, Goldsmiths, University of London, UK

Melodic discrimination has formed the basis of much research, both in the experimental psychology tradition (e.g. Dowling, 1971) and in the individual differences tradition (e.g. Seashore, 1919). In the typical melodic discrimination paradigm, participants are tasked with detecting differences between two successively played versions of the same melody. Experimental psychology studies have investigated how melodic features influence task performance, identifying effects from a range of sources including contour, length, and tonal structure. Meanwhile, individual difference studies have found that melodic discrimination ability provides a good proxy for musical expertise, and as a result the task is often used in musical test batteries (Law & Zentner, 2012). However, there is surprisingly little consensus concerning what cognitive ability or abilities the melodic discrimination test actually measures, raising serious problems for its construct validity. The present research aimed to address these problems by forming and testing an explicit cognitive model of melodic discrimination. This cognitive model comprises four main stages: perceptual encoding, memory retention, similarity comparison, and decision-making. On the basis of this model, we hypothesised that item difficulty should be predicted by the complexity of the reference melody (i.e. first version in the pair), the tonalness of the reference melody, and the perceptual similarity between melody versions. We operationalise these three features using computational models (SIMILE, Müllensiefen & Frieler, 2003; FANTASTIC; Müllensiefen, 2009). We then test the influence of these three features on melodic discrimination performance by applying explanatory item response modelling to response data collected for three pre-existing melodic discrimination tests (N = 317). Perceptual similarity and complexity measures significantly predicted task difficulty, supporting the proposed four-stage cognitive model and contributing to the construct validity of the melodic discrimination test.

Measuring and Modeling Melodic Expectation: A New Approach

Allison R. Fogel,*[1] Emily Morgan,*[2] Aniruddh D. Patel*[3]

* Department of Psychology, Tufts University, Medford, MA, USA
[1]Allison.Fogel@tufts.edu, [2]Emily.Morgan@tufts.edu, [3]a.patel@tufts.edu

ABSTRACT

Melodic expectancy has been a focus for quantitative modeling within music cognition. We recently developed a melodic cloze probability task (modeled on the well-known linguistic cloze task) in which listeners are presented with the beginnings of novel tonal melodies and are asked to sing the note they expect to come next. In our initial study, half of the melodies had an underlying harmonic structure designed to constrain expectations for the next note, while half did not. As predicted, participants' responses were significantly more consistent following constraining melodies. Here, we compare our participants' expectations to the simplified version of the implication-realization (IR) model of melodic expectancy. When the model was fit to the combined data from both conditions, it did not distinguish between the AC and NC experimental conditions, underpredicting tonic responses in the AC condition and overpredicting tonic responses in the NC condition. Overall, our results indicate that the principles of the IR model do significantly affect participants' responses, but that they still leave much of the variance in participants' responses unexplained. In future work, we plan to expand upon these regression analyses to explicitly compare different quantitative models of melodic expectancy.

I. INTRODUCTION

Different forms of musical expectancy (such as melodic, rhythmic, and harmonic expectations) have been the subject of many theoretical and empirical investigations in music cognition (e.g., Schmuckler, 1989; Narmour, 1990; Krumhansl et al., 1999; Large and Jones, 1999; Huron, 2006). Melodic expectancy in particular has been a focus for quantitative modeling (e.g., Schellenberg, 1996; Krumhansl et al., 1999; Margulis, 2005; Pearce et al., 2010).

Two open questions in the study of melodic expectancy are to what extent expectations depend on the local context (e.g. the preceding one or two notes) versus the larger context of a melody, and to what extent these expectations are shaped by statistical learning versus by innate, broadly applicable psychological/perceptual principles. Narmour's (1990) implication-realization (IR) model of melodic expectancy, which we focus on in this paper, claims that expectations are shaped by a small number of factors grounded in human principles of perceptual organization, similar to Gestalt principles (such as a prediction for successive notes to be proximate in pitch to previous notes). Although these principles can in theory be applied at different levels of musical structure, the quantitative version of this model developed by Schellenberg (1997) focuses on local note-to-note relationships.

Other models provide different answers to the above questions. For example, Pearce's (2005) IDyOM model also focuses on the role of local context in expectancy, but claims that expectations are learned from statistical tendencies in both one's previous musical experience and within the current melody. Koelsch et al. (2013) argue that expectations are determined not only by local context but by language-like hierarchical structure.

Here we present a test of Narmour's IR model using a novel musical production task. While melodic expectancy has been measured using several different behavioral tasks in the past, these studies have generally focused on perception rather than production. We recently developed an approach to studying melodic expectation that allows fine-grained control over melodic structure and fine-grained measurements of expectation with a naturalistic, online behavioral task. In the melodic cloze probability task (modeled on the widely used linguistic cloze probability task), listeners are presented with the beginnings of novel tonal melodies and are asked to sing the note they expect to come next (Fogel et al., 2015). For a given melody, the percentage of participants providing a given continuation is taken as the "cloze probability" of that response.

Here, we compare participants' expectations to the simplified version of the well-known implication-realization (IR) model of melodic expectancy (Narmour, 1990) developed by Schellenberg (1997). This model uses two melodic factors to predict the probability of each possible continuation. "Pitch proximity" refers to listeners' expectations that the next tone of a melody will be proximate in pitch to the last tone heard. "Pitch reversal" states that after a leap, listeners anticipate that the next tone will reverse direction and land in a pitch region proximate to the penultimate tone.

II. METHOD

In our initial study, melodies were composed in 45 pairs and then truncated to create "melodic stems" of 5-9 notes. One melodic stem in each pair had an underlying harmonic structure designed to constrain expectations for the next note to a specific continuation by ending with an implied authentic cadence (AC condition); the other was matched in terms of length, rhythm, melodic contour, and key, but ended with an implied IV, vi, or ii harmony and was not designed to constrain expectations (NC condition; see Figure 1).

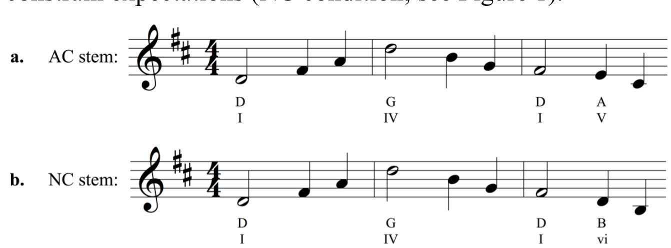



Figure 1. (A) Authentic cadence (AC) and (B) Non-cadence (NC) versions of one melodic pair (see text for explanation). The figure shows the AC and NC stems in Western music notation. Shown beneath each stem is a possible interpretation of the underlying implied harmonic progression (e.g., D, G, D, A chords in the AC stem), and harmonic functions (I, or tonic chord; IV, or subdominant chord; V, or dominant chord; vi or submediant chord).

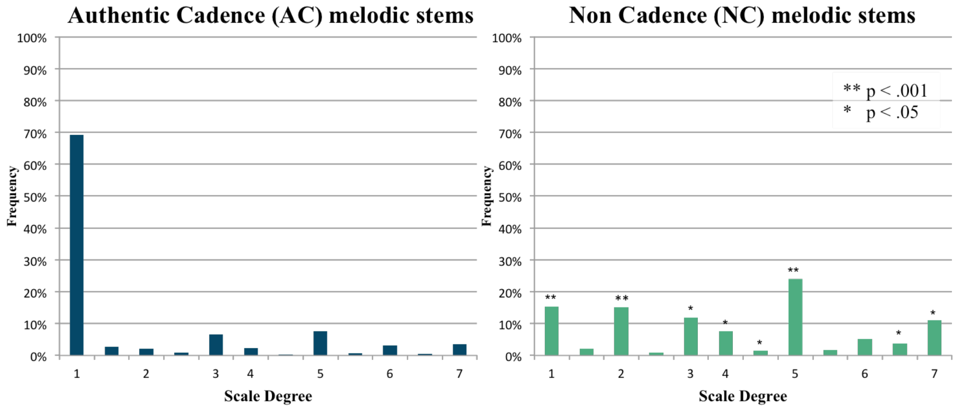


Figure 2. Average of all response distributions to AC and NC stems, shown as scale degrees. Numbers represent diatonic (major) scale degrees (e.g., 1 = tonic, 7 = leading tone, etc.), with asterisks indicating scale degrees with significantly different frequencies between the two conditions.

Participants were 50 self-identified musicians with a minimum of 5 years of musical training within the past 10 years and a mean of 9.0 years of formal musical training on Western musical instruments. Participants were instructed that they would hear the beginning of a melody and would need to "sing the note you think comes next" to continue (not necessarily complete) the melody. Each participant was presented with 24 AC and 24 NC melodic stems (only one version from each melodic stem pair; three pairs were eliminated from further analysis). Responses were recorded and the mean fundamental frequency of each sung response was rounded to the closest semitone.

In previous work (Fogel et al., 2015), we found that, as predicted, participants' responses were significantly more consistent following AC stems, where an average of 69% of participants sang the same note. In contrast, after NC stems, an average of only 42% of participants sang the same note. When responses are represented as diatonic scale degrees, responses to AC stems were highly weighted around the tonic (or first note of the scale), while responses to NC stems were more widely distributed across in-key scale degrees (see Figure 2).

In the current work, we ask to what extent participants' responses are predicted by the IR model of melodic expectancy.

III. RESULTS

A. Data Analysis

We use regression analyses to determine to what extent the two predictors of the IR model affect participants' responses. Specifically, we analyze the data using multinomial discrete-choice logit models, implemented using the mlogit package in R (Croissant, 2013). For simplicity, only responses of in-key notes are considered, excluding 7.6% of responses that consisted of out-of-key notes. Responses are coded as scale degrees relative to the key of the melody and treated as unordered categories, with the tonic used as the reference level.

The independent variables considered are Proximity and Pitch Reversal, coded numerically as described by Schellenberg (1997) and implemented in the MIDI toolbox (Eerola & Toiviainen, 2004). In particular, Proximity is coded as distance in semitones from the final note heard to the response note, such that the predicted effect of this predictor is negative: i.e. we predict that response notes at higher distances have lower probability of being chosen. The predicted effect of Pitch Reversal is positive, with higher values indicating better fit according to Schellenberg's model. An intercept term is also included in the model to account for differing base rates for the response categories.

Random by-subjects slopes for Proximity and Pitch Reversal are included, following Barr et al. (2013). However, no by-item random effects are included as mlogit—and in fact all other standard multinomial discrete-choice logit model-fitting software as well—is unable to account for crossed random effects. Thus, p-value estimates obtained from our model may be anti-conservative as item-level effects are unaccounted for. We plan to address this in future analyses.

B. Results: Both conditions

A realistic model of melodic expectancy must account for cases in which a cadence occurs and cases in which one doesn't, i.e. it should work equally well in both experimental conditions. Thus, to provide the strongest test of our model, we fit it on the combined data from both conditions. The model reveals significant effect of both Proximity (β = -0.20, t = -16.06, p<0.001) and Pitch Reversal (β = 0.20, t = 5.61, p<0.001) in the predicted directions, indicating that these two factors play a significant role in determining participants' responses as predicted by the IR model. McFadden's R^2, an approximation of a traditional R^2 value for multinomial logit models, for this model is 0.041, indicating that the current model explains relatively little of the variance in the data. (For models of trial-level data such as this, values of McFadden's R^2 in the 0.2-0.4 range are considered excellent; McFadden, 1978).

We can use this model to predict the distribution of responses for each melody and compare this against the true empirical distribution. The mean L1 error across all notes in all melodies (i.e. the average difference between predicted and empirical probabilities) is 0.11 (sd=0.035).

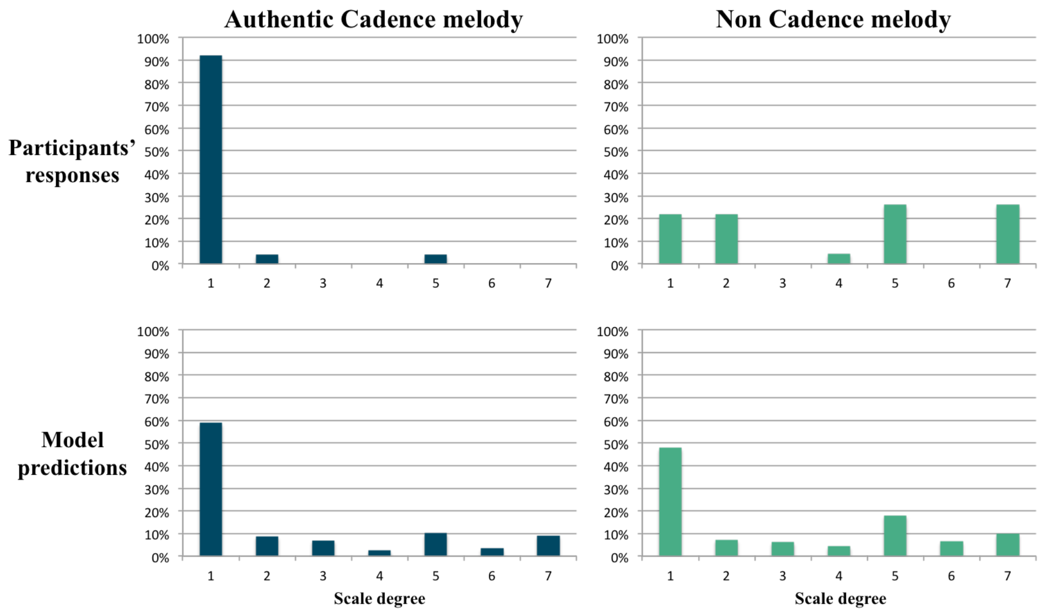


Figure 3. Comparison of the true empirical distribution of participants' responses to the melody pictured in Figure 1 to the distribution predicted by the IR model. The model does not do well in distinguishing between the AC and NC experimental conditions, underpredicting tonic responses in the AC condition and overpredicting tonic responses in the NC condition.

As shown in Figure 3, although the IR model is capturing *some* of the variance in the data, it leaves much unexplained. In particular, it does a poor job of distinguishing between the AC and NC experimental conditions, underpredicting tonic responses in the AC condition and overpredicting tonic responses in the NC condition. To further explore the predictions of the IR model in these two conditions, we re-fit the multinomial discrete-choice model on data from each condition separately.

C. Results: Non Cadence condition

Results of the model fit to NC condition responses reveal significant effects of both Proximity (β = -0.23, t = -12.85, p<0.001) and Pitch Reversal (β = 0.14, t = 3.02, p<0.01) in the predicted direction. McFadden's R^2 for this model is 0.047. As before we use the fitted model to predict the distribution of responses for each melody and compare it against the true empirical distribution. The mean L1 error is 0.099 (sd=0.032). On all these measures, the NC model appears comparable to the both-conditions model.

D. Results: Authentic Cadence condition

Results of the model fit to AC condition responses reveal a significant effect of Proximity (β = -0.15, t = -6.95, p<0.001) in the predicted direction and a significant effect of Pitch Reversal *opposite* the predicted direction (β = -0.19, t = -2.44, p=0.01). McFadden's R^2 for this model is 0.038. The mean L1 error is 0.050 (sd=0.023).

Although the absolute (L1) error is lower for the AC model than for the NC or both-conditions models (indicating a better fit between predictions and true responses) McFadden's R^2 is also lower, indicating that *less* variance in the data is explained by this model. This paradox reflects the fact that the response data itself is less variable in the AC condition, as responses in this condition predominantly consist of the tonic. The AC model has better absolute performance by virtue of learning an intercept term that strongly favors always choosing the tonic, but the remaining variance is not well accounted for. Moreover, note that the model's ability to learn an intercept term that favors the tonic is independent of the IR model's predictions. Thus it appears that the IR model itself is unable to account for the fact that the tonic is strongly favored in these melodies.

IV. CONCLUSION

Our results indicate that the principles of the IR model do significantly affect participants' responses but that they still leave much of the variance in participants' responses unexplained. The Proximity factor in particular is reliably significant in the predicted direction, while we found mixed results for the Pitch Reversal factor.

A comparison of the AC and NC melodies revealed that the IR model is unable to predict the strong preference for tonic completions in the AC melodies. In other words, the note-to-note principles encoded by this model are unable to capture the underlying harmonic structure that generates strong expectations for an authentic cadence.

In future work, we plan to expand upon these regression analyses to explicitly compare different quantitative models of melodic expectancy such as the IDyOM model (Pearce, 2005) and hierarchical models (Koelsch et al., 2013), which—by taking into account more (and different) sources of context—may better be able to account for expectations such as those in the AC melodies. Predictions of multiple models can be entered into the same regression model in order to test which better predicts the given data. This will allow us to directly test whether models focusing on local versus global context, or learning statistical relationships versus innate principles, are better predictors of participants' responses.

REFERENCES

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278.

Croissant, Y (2013). *mlogit: multinomial logit model*. R package version 0.2-4. URL https://cran.r-project.org/web/packages/mlogit/index.html.

Eerola, T., and Toiviainen, P. (2004). *MIDI Toolbox: MATLAB Tools for Music Research*. Jyväskylä: University of Jyväskylä.

Fogel, A. R., Rosenberg, J. C., Lehman, F. M., Kuperberg, G. R., & Patel, A. D. (2015). Studying Musical and Linguistic Prediction in Comparable Ways: The Melodic Cloze Probability Method. *Frontiers in Psychology*, *6*:1718.

Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: MIT Press.

Koelsch, S., Rohrmeier, M., Torrecuso, R., & Jentschke, S. (2013). Processing of hierarchical syntactic structure in music. *Proceedings of the National Academy of Sciences of the United States of America*, 110(38), 15443-15448.

Krumhansl, C. L., Louhivuori, J., Toiviainen, P., Järvinen, T., & Eerola, T. (1999). Melodic expectation in Finnish spiritual folk hymns: Convergence of statistical, behavioral, and computational approaches. *Music Perception*, 17, 151-195.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, *106*, 119.

Margulis, E. H. (2005). A model of melodic expectation. *Music Perception*, *22*(4), 663- 714.

McFadden, D. (1978). Quantitative Methods for Analysing Travel Behaviour of Individuals: Some Recent Developments. In D. A. Hensher & P. R. Stopher (Ed.), *Behavioural Travel Modeling* (pp. 279-318). London: Croom Helm.

Narmour, E. (1990). *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model*. Chicago: University of Chicago Press.

Pearce, M. T. (2005). *The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition*. Ph.D. Thesis, City University, London.

Pearce, M. T., Ruiz, M. H., Kapasi, S., Wiggins, G. A., & Bhattacharya, J. (2010). Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation. *NeuroImage*, *50*(1), 302-313.

Schellenberg, E. G. (1996). Expectancy in melody: Tests of the implication- realization model. *Cognition*, *58*(1), 75-125.

Schellenberg, E. G. (1997). Simplifying the implication-realization model of melodic expectancy. *Music Perception*, *14*(3), 295-318.

Schmuckler, M. A. (1989). Expectation in music: investigation of melodic and harmonic processes. *Music Perception, 7,* 109–149.

Musical Stylometry, Machine Learning, and Attribution Studies: A Semi-Supervised Approach to the Works of Josquin

Andrew Brinkman*[1] Daniel Shanahan, *[2] Craig Sapp#[3]

*School of Music, Louisiana State University, Baton Rouge, Louisiana, USA
#CCARH/PHI, Stanford University
[1]abrink4@lsu.edu, [2]dshanahan@lsu.edu, [3]craig@ccrma.stanford.edu

ABSTRACT

Compositional authorship is often assigned though factors external to the musical text, such as biographical records and surveys of source attributions; however, such methodologies often fall short and are potentially unreliable. On the other hand, determining compositional authorship through internal factors—through stylistic traits of composers derived from the music itself—is often fraught with errors and biases. One of the underlying assumptions in the field of stylometry is that, while it is difficult for humans to perform a truly unbiased analysis of authorial attribution, computational methods can provide clearer and more objective guidelines than would otherwise be apparent to readers or listeners, and thus might provide corroboration or clues for further investigation. This paper discusses machine-learning approaches for evaluating attribution for compositions by Josquin des Prez. We explore musical characteristics such as melodic sequences, counterpoint motion, rhythmic variability, and other entry measures to search for features inherent to a composer's works or style, and we hope that employing such an approach—one that explicitly states which factors led to the decision-making process—can serve to inform scholars looking at other works and composers.

I. INTRODUCTION

Perhaps more than any other composer, Josquin has generated the most scholarly debate regarding the attribution of his works: those that were at one point seen as firmly rooted in the composer's *oeuvre* are now frequently discredited and are being reconsidered. This controversy stems from a number of factors, including the lack of primary source material. In the case of the *New Josquin Edition*, a collection of Josquin's music sorted by various factors (including controversial attribution levels), it often designates a work's authorship by the amount of times that work appears within primary collections (Wegman, 2008).

Only a few other factors contribute to determining attribution within these works. These primarily include whether or not a work appears in direct reference to the composer in biographical accounts of the time. Additionally, of the over 400 works attributed to the composer across multiple genres, more than 200 of them are considered to have questionable levels of attribution (Rodin and Sapp, 2016). This, coupled with uncertainties about sources for printed editions of his works (particularly for posthumous publications), increases the likelihood of incorrect attribution. Outside of the appearance of works within primary sources, there is no consistently reliable process for determining attribution. While many musicologists have suggested a variety of methods to help determine authorship without the use of primary sources, these methods are unreliable and frequently contested (see Sparks 1971; Wegman, 2008, and Jas, 2014). The Josquin Research Project (JRP) has the most rigorous and systematic in its approach to identifying possibly problematic attributions, doing so based solely on early manuscript sources. In the JRP, an attribution level of "1" means the piece is attributed to the composer in at least two early manuscripts, whereas a piece with an attribution level of "2" is referred to in only one early source. Levels 3–6 represent decreasing levels of certainty about pieces for which there are no early manuscript sources.[1]

Stylistic features are debated between scholars as the composer's style may have likely changed over the course of their career, and what were once thought to be idiosyncrasies are often shown to appear more broadly in the repertoire. This study tries to mitigate this issue through multiple machine-learning models. By utilizing both high-level features suggested by these musicologists as well as several low-level features (such as note-to-note transitions and correlations to securely-attributed works), we hope to be able to provide more likely starting points for researchers, as well as a model that will ideally grow into a tool for more general attribution studies in music research.

II. MUSICAL STYLOMETRY

Stylometry has been used in the humanities to better determine and understand the authorship of works ranging from the *Federalist Papers* (Motseller & Wallace, 1964) to the works of Shakespeare (Merriam & Matthews, 1994), and more recently the pseudonymous work of J.K. Rowling (Juola, 2015). Similar statistical approaches have even been used to analyze what might be coerced confessions from suspects (and possibly forged confessions by law enforcement) in criminal trials (Kenny, 2013). A central principle of computational stylometry is that the low-level features that are often undetectable to even the most experienced reader can serve as a sort of "authorial fingerprint", distinguishing an author from their contemporaries and colleagues (see Love, 2002).

This method is viewed—often justly—with a certain level of skepticism. Early approaches had many results that were eventually disproven, such as A.Q. Morton's analysis of the works of Joyce (see Levitt, 2001). Recent models have improved with advances in statistical methods and machine-learning techniques. Models tend to use a mixture of high-level features, such as word preferences (e.g. the universal "he", "she", "he/she", or the singular "they") and intentional coinages, such as those employed by Coleridge (see Love, 2002; pp.107–108), as well as low-level features, such as the

[1] for more information, see http://josquin.stanford.edu/about/attribution.

transitional probabilities of certain words, and the correlation of the word frequencies to a corpus of confirmed authorship. Music researchers have occasionally used such techniques to determine compositional authorship. For example, Backer and Kranenburg (2005) used a linear discriminant transformation to examine the features of works by J.S. Bach and J.L. Krebs (see also Kranenburg, 2007), and Bellman (2011) employs computational methods to quantify musical style engaging with stylometric questions throughout his study. Speiser and Gupta (2013) use principal components analysis to train classifiers (similarly to how this paper approaches the problem). The current study makes use of these previously implemented approaches to better understand the authorship of the works of Josquin.

A. High-Level Musical Features

This study makes use of what we might consider "high-level" features, meant to accompany the currently employed "low-level" features. The former category includes the presence of 9-8 suspensions, as well as the rate of occurrence of the specific types of contrapuntal motion: parallel, similar, contrary, and oblique. Many other commonly-discussed high-level features are not employed in the current study. For example, musicologists often debate the uniqueness of the texture around cadence points in Josquin's music, arguing that textures become sparser as cadences approach (Wegman, 2008). Such features can be difficult to quantify, and therefore difficult to operationalize. Other features commonly discussed but left out of the current study include "relaxed" contrapuntal writing and an overreliance on repetition of material. These both require a level of interpretation that we decide not to employ for the current study.

Additionally, a feature commonly referred to as the *Satzfehler* has been discussed and debated. Originally coined by Helmuth Osthoff (Osthoff, 1962), the *Satzfehler* was considered to be a possible error in voice-leading made by Josquin during his early years, in which leading-tone motion occurred between two voices simultaneously and involved one voice dipping down from the tonic to the leading-tone and then ascending back to the tonic while another voice preemptively performs the leading-tone before the first voice has had a chance to resolve. Later Edgar Sparks argued that this device was not an error but was simply a common feature used, not only by Josquin, but also by many of his contemporaries (Sparks, 1976). While the appropriateness of this technique is still being questioned, it is one of the few techniques that have explicitly been applied to Josquin's music as an identifier of his compositional style, and have been documented at length (Wegman, 2008). For the current study, however, we decided to exclude it from our searching, as once again there were many flexibilities to the definition, and computational searching yields too many false positives.

B. Low-Level Musical Features

For low-level features, we examined note-to-note transitions by scale degree. Admittedly, the notion of scale degree does not easily map onto the music of the Renaissance, but it nevertheless provides a concise method for looking at relationships between notes. These were then examined along with other similar low-level features including rhythmic variability, measured with the normalized pairwise variability index (see Grabe and Low, 2002; Daniele and Patel, 2004; Patel and Daniele, 2013), the average entropy of each melodic voice, and the correlation to the pitch distributions in secure Josquin works.

Table 1. Features contributing to the models.

High-Level	Low-Level	Metadata
9-8 suspensions	average melodic entropy (per voice)	primary listed composer
oblique motion	nPVI	title
contrary motion	pitch-range correlation to secure Josquin works	JRP attribution level
similar motion	note-to-note transition probabilities (by scale degree)	genre
parallel motion		

III. BETWEEN-COMPOSER COMPARISON

We compare the music of Josquin to that of other composers in a range of styles. This helps to identify features that represent differences between styles, are common to a particular style, or are unique to a particular composer. A comparison between Josquin and J.S. Bach's four-part chorales is used as an initial calibration comparison. Most any untrained listener should be able to distinguish between these two repertories separated by two centuries. Slightly more difficult are comparisons to Ockeghem and Du Fay, who predate Josquin by one and two generations. These comparisons are more difficult for the average listener, but relatively easy for experienced listeners of Renaissance music. Finally, challenging comparisons are made to de Orto, and La Rue. These composers are Josquin contemporaries who all compose in a common style, and their works are often misattributed amongst each other.

A. Dimensionality Reduction

Before training a classifier, we employed principal component analysis (PCA) to reduce the 53 input features into a more manageable number of dimensions. This maintains as much variance between features as possible and is used to drop dimensions containing little information. High-level feature comparisons seem to be more salient between genres and epochs, whereas low-level features function best within them. To anticipate the results of this portion of the study, comparisons between disparate styles (Josquin and Bach) were actually hindered by the inclusion of low-level features, and the classifiers performed better when only provided the high-level information. In contrast, low-level features do increase the separation abilities of PCA and can train more accurate classifiers when provided with whose music in style similar to that of Josquin.

B. Josquin versus Bach

Before training and testing a model on composers with similar styles to Josquin, we examine differences between the works of Josquin and Bach chorales. If musical features were to have difficulties between two composers, then it would be logical to assume that progression to more similar style comparisons would prove to be problematic, let alone be suitable for authorship studies. High-level features were used to compare the works of Josquin (those listed as secure according to JRP attribution assignments) and the Bach chorales available in kern format.[2] When applied to only the high-level features, PCA shows that most of the variance can be accounted for within the first two principal components. As can be seen in Figure 1, the elbow might argue that the model would be feasible with only these two components. We instead opted to include the first five components, which account for 85% of the variance.

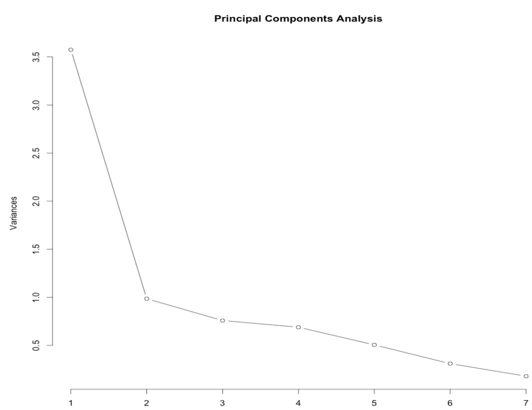


Figure 1. Variance accounted for in each principal component when comparing Bach and Josquin.

Figure 2 illustrates the strength of each high-level feature in this model. The largest amount of variance seems to be accounted for by the contrapuntal and melodic techniques (9-8 suspensions, similar, parallel, and oblique motion), as well as rhythmic variability (the nPVI of the entire piece, notated as nPVI_entire). This will makes sense to anyone who has spent any amount of time listening to these two composers.

A relatively clear separation can be seen in Figure 3, with the model able to discern the music of Bach from Josquin's music quite easily and from relatively few features. Oblique motion, pitch entry and pitch correlations are the best base features to distinguish between these two repertories. As we will see with our classifiers, these principal components can be used to train a model that can easily discern Bach from others composers being evaluated (Table 2).

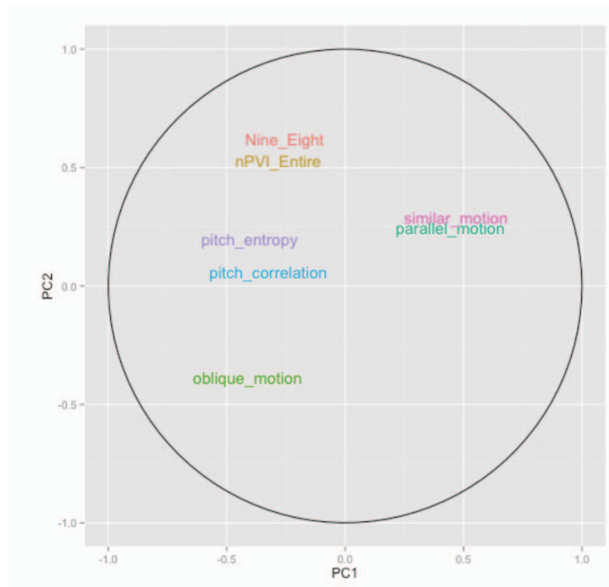


Figure 2. PC coefficients associated with the variables between the Josquin and Bach datasets.

Figure 3. The separation of Bach and Josquin given only the two PCs that account for the most amount of variance (full model uses 5 PCs).

C. Josquin and Du Fay

A more difficult task involves comparisons to works by composers closer to Josquin's lifetime (*c*1450–1521). Guillaume Du Fay (1397–1474) provides a good starting point, as the two are separated by two generations. Low-level features were included for this comparison to facilitate separation. Figure 4 shows the contribution of features to the first two principal components. Repetition of the fifth scale degree seems to account for the highest level of variance between these two composers as seen by the "t5_5" label in the figure with a correlation around 0.25 for the first principal components axis.

[2] https://github.com/craigsapp/bach-371-chorales. JRP scores can most conveniently also be downloaded from Github at https://github.com/josquin-research-project/jrp-scores.

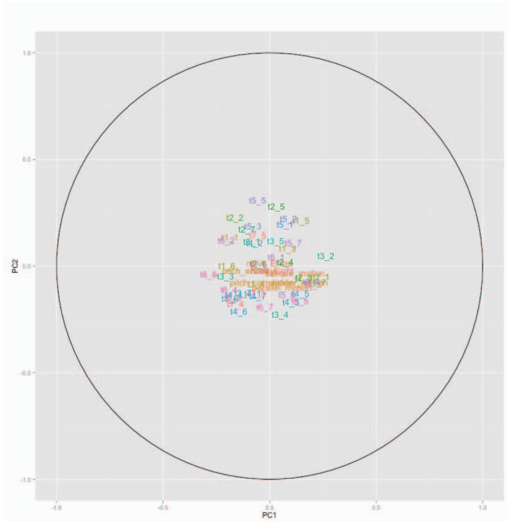


Figure 4. PC coefficients associated with the variables between the Josquin and Du Fay datasets.

Looking at the biplot in Figure 5 for the separation, we see that the first two PCs are not able to separate the two composers quite as well as with Josquin/Bach. The model relies on more features (the features in addition to the high-level), so each PC accounts for less variance.

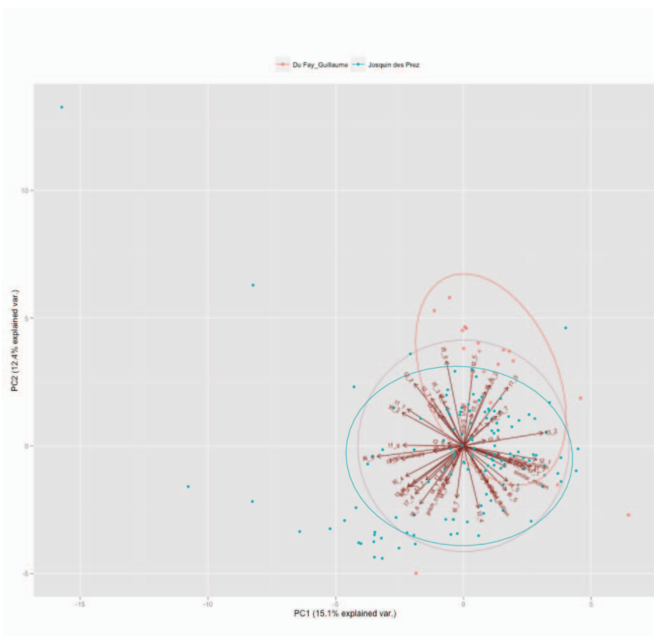


Figure 5. The separation of Josquin and Du Fay visualized using only the first two PCs.

D. Josquin and Ockeghem

Johannes Ockeghem's (*c*1425–1497) style is yet more similar to Josquin's than Du Fay or Bach. Yet expert listeners can distinguish relatively easily between the styles of Josquin and Ockeghem. As with Du Fay, repetition of the fifth stands is a distinguishing feature between the two composers, and (unlike Du Fay) the treatment of the second scale degree accounts for a decent amount of the variance in the top two principal components.

We might predict that our classifiers would score fairly well in classifying between these two composers, although less well than Bach, and perhaps less clearly than Du Fay.

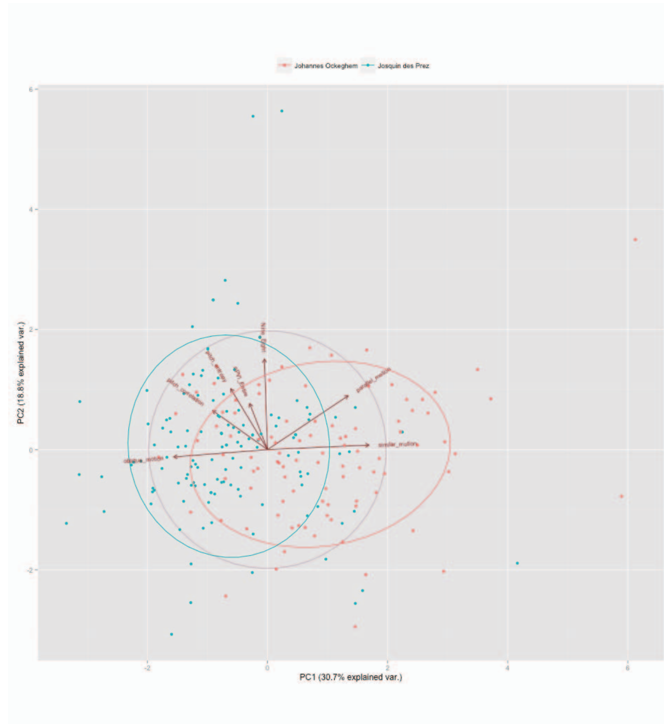


Figure 6. The separation of Josquin and Ockeghem visualized using only the first two PCs.

E. Josquin and de Orto

Mabrianus de Orto (1469–1529) is more closely a contemporary of Josquin's, and their styles even closer than the previously mentioned composers. Interestingly, the repetition of the fifth continues to play a strong role, but with a negative correlation. Melodic sequences involving the fourth (3 → 4 and 4 → 6) strongly correlate between the two composers. It is likely that the model would perform fairly poorly on de Orto, and this might be hinted at in Figure 7's biplot. In this case the first PCs describing de Orto's works is nearly completely encompassed within that of Josquin's works.

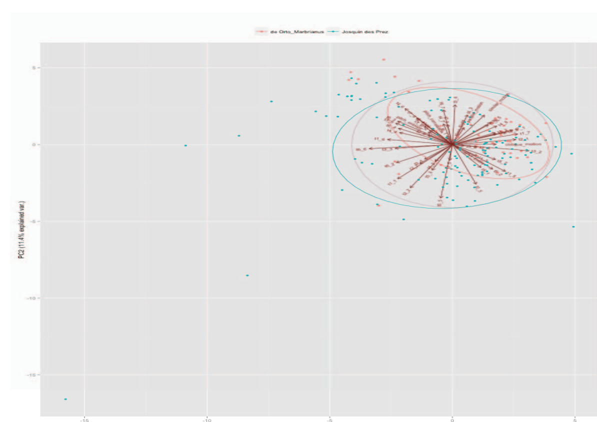


Figure 7. The separation of Josquin and de Orto visualized using only the first two PCs.

F. Josquin and La Rue

Pierre de la Rue (c1452–1518) is more problematic than de Orto. His works are most commonly mistaken for Josquin's (and vice versa), and the two are the most stylistically similar of all of the composer's examined in the present study. Figure 8, shows an interesting pattern: the repetition of the fifth is anti-correlated, and the strongest correlation coefficients are provided by 3 → 4 scale degree motions.

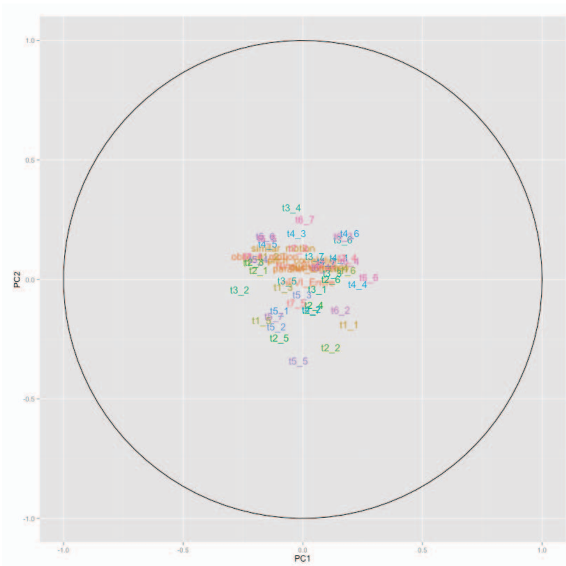


Figure 8. First two PC coefficients for Josquin and La Rue features.

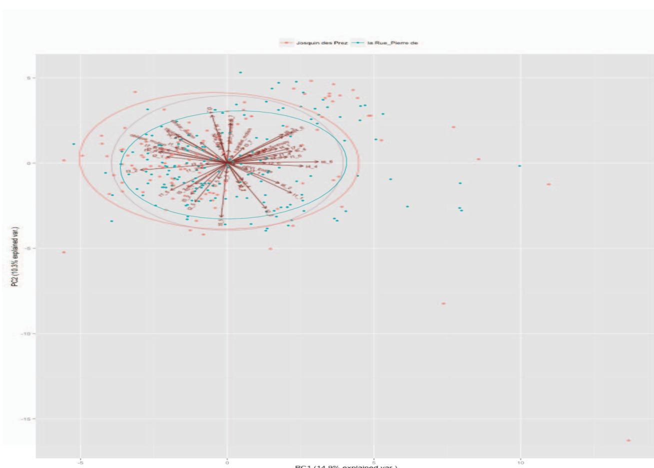


Figure 9. The separation of Josquin and La Rue visualized using only the first two PCs.

Figure 9 shows the biplot for Josquin and La Rue. The two composers complete overlap each other (and thus the first two PCs cannot be used alone to distinguish between the composers at all). This figure illustrates why many researchers have mistakenly cross-attributed the works of both composers.

IV. CLASSIFIER PERFORMANCE

While the previous section focused on binary comparisons between Josquin and another composer, attribution does not exist in such tidy laboratory conditions. Ideally, all composers should be fed into a classifier. In order to achieve this goal, the current study uses the results of the principal component analysis as a way of training a multi-composer classifier. As this study is primarily exploratory in nature, we opted to train three types of classifiers, and compare the results. A future goal, as will be discussed below, would be to employ an ensemble-learning model, in which multiple models are used and a decision tree is implemented to take the results of the model that best performs in a given context. The current study employs a k-nearest neighbor (KNN), a support vector machine, and a decision tree (as a way of better understanding the role of each feature).

A. K-Nearest Neighbor Classifier

A principal component analysis was performed on all of the above composers at once, using 48 features. In order to account for 85% of the variance (the standard used above), 27 PCs were needed to create the model. A two-dimensional representation of the separation between the first two PCs can be seen below in Figure 10. Notice that the first two PCs can be used to anticipate the similarity of style between the composers. Bach (red) is well separated from all of the Renaissance composers, while Du Fay (green) is the second more differentiated. The first two PCs cannot differentiate by themselves between the other composers' music (Ockeghem, de Orto, La Rue and Josquin).

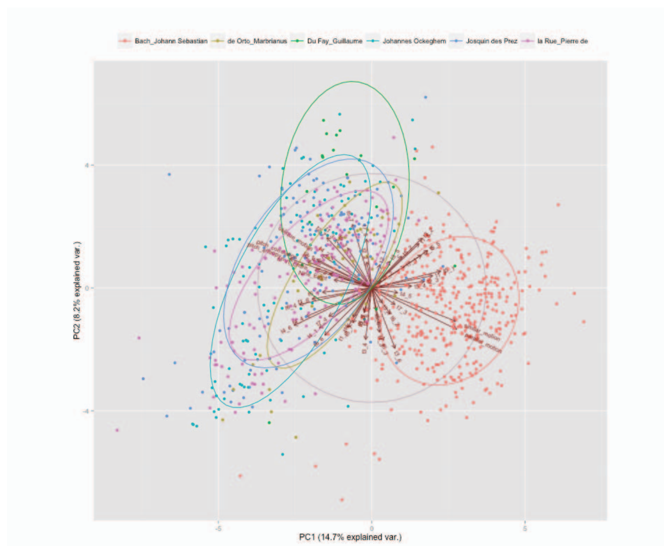


Figure 10. The separation of all composers visualized using only the first two PCs (this model incorporated 27 principal components).

An 80/20 split was used (the model was trained on 80% of the data and tested on 20%), with k=6. A confusion matrix is provided below in Table 2. Overall, the model performed quite well in discerning Bach from the other composers, indicating once again that very little is needed in order to determine pieces by genre and time-period. One Bach chorale was mistakenly assigned to de Orto, and another to Josquin. The model correctly predicted de Orto 38.9% of the time, but assigned a couple of pieces each to Ockeghem and Josquin. The model's identification of de Orto's works was most confused by La Rue, mistakenly assigning de Orto pieces to him a third of the time.

Du Fay was most confused with Ockeghem, which makes sense, given their relative proximity in time period and similar compositional styles. Interestingly, however, this was not a bidirectional confusion: Ockeghem pieces were rarely confused for those written by Du Fay, and instead were more likely to be ascribed to La Rue (20% of the time). The model did surprisingly well with discerning Josquin from La Rue, which was predicted to be the most difficult comparison. Although 21.2% of Josquin pieces were confused for La Rue pieces, the model guessed correctly on 60.6% of the pieces. For La Rue, the model was accurate 80.6% of the time, and only ascribed pieces to Josquin 16.7% of the time. While, at first sight some of these accuracy levels might seem to be a little low—we had hoped for at least a 75% accuracy with Josquin—it should be noted that each composer performed significantly better than the baseline (which would be around 16.7%, given a uniform random comparison between six composers).

Table 2. Confusion matrix for KNN model. Results are in a percentage of responses. The correct answer is listed in the first column, and predictions are listed in the top row.

	Bach	de Orto	Du Fay	Ock.	Josquin	La Rue
Bach	94.5	.9	0	0	.9	3.6
de Orto	5.6	38.9	0	11.1	11.1	33.3
Du Fay	14.3	0	42.9	28.6	0	14.3
Ockeghem	0	6.7	3.3	70	0	20
Josquin	0	9	3	6.1	60.6	21.2
La Rue	0	2.8	0	0	16.7	80.6

B. Support Vector Machine

We next trained a support vector machine (SVM) classifier, intended to improve on the results of the first. KNN have nice properties such as automatic adaption to non-linear data, but it has to be tuned carefully and is sensitive to bad features. SVMs are useful for segmenting data in high-dimensional spaces. The SVM used a radial basis function kernel which allows for a classification boundaries that might curved, rather than linear. The same PCs were used as in the k-nearest neighbor classifier.

The results, which can be seen in Table 3, are limited in scope: there was a limited number of pieces in the random test set available for composers such as de Orto and Du Fay. Nevertheless, the results were consistent across multiple tests. Interestingly, the model does not seem to do much better when predicting Josquin, but the accuracy improves slightly with La Rue, while the Bach prediction remains quite high. Surprisingly, the SVM performs similarly to the KNN method, with ratings close to the assignments in Table 2.

Table 3. Confusion matrix for SVM using a radial basis function. Results are in percentage of responses. The correct answer is listed in first column, top row lists predicted composer.

	Bach	de Orto	Du Fay	Ock.	Josquin	La Rue
Bach	98.5	0	0	0	0	1.5
de Orto	33.3	33.3	0	0	0	33.3
Du Fay	50	0	25	25	0	0
Ockeghem	6.7	0	0	60	13.3	20
Josquin	16	0	4	4	60	16
La Rue	0	0	0	0	12.9	87.1

C. Future Work

Future work will examine the specifics of each feature more thoroughly as a way of identifying the differences between composers. For example, the features of all of the composers studied were placed within a decision tree model to better understand how such aspects might contribute to creating better models in the future. As can be seen below, the most important features in discerning the four composers seem to be the employment of similar motion, as well as treatment of the sixth scale degree, followed by motion of scale degree two to scale degree five. While the decision tree below divides the pieces into eight, rather than six, compositional styles, it allows us to better understand the differences the constitute each composer's style and provide some context for the above PCA and subsequent classifiers.

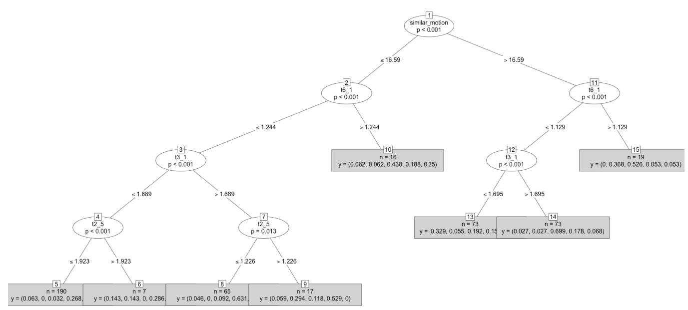


Figure 11. Decision tree analysis diagram

V. CONCLUSION

The current study attempts to demonstrate the applicability of machine-learning techniques to the question of authorship in Josquin's music. There are obviously quite a few points that need to be further discussed. For example, genre matters. Composers wrote differently for masses than they did for motets, yet the current study (primarily to have as much statistical power as possible) treats each composer's output as a uniform stylistic output. Secondly, it assumes that compositional style is undergoes minimal levels of change throughout the course of a composer's career. Future work hopes to be able to use these elements, although the sources are often sparse and apocryphal. Additionally, we hope to examine more features in future work, including harmonic entropy, rhythmic transition probabilities, and the incorporation of the aforementioned *Satzfehler*. Once these are implemented we hope to provide a more accurate model that can serve as a tool for researchers investigating this music.

ACKNOWLEDGMENTS

We would like to thank Jesse Rodin and the entire team at the Josquin Research Project for all of the work they do on the project, and Blake Howe for his helpful advice and insight into the music of Josquin.

REFERENCES

Backer, E., & Kranenburg, P. van. (2005). On musical stylometry—a pattern recognition approach. *Pattern Recognition Letters*, 26(3), 299–309.

Bharucha, J. J. (1991). Pitch, harmony, and neural nets: A psychological perspective. In P. Todd & G. Loy (Eds.), *Music and Connectionism* (pp. 84–99). Cambridge, MA: MIT Press.

Daniele, J. R., & Patel, A. D. (2013). An Empirical Study of Historical Patterns in Musical Rhythm: Analysis of German & Italian Classical Music Using the nPVI Equation. *Music Perception*: *An Interdisciplinary Journal*, (31/1). pp. 10–18.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the Rhythm Class Hypothesis. In Laboratory Phonology, (7). pp. 515–546. Berlin: Mouton de Gruyter.

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. Cambridge, MA: MIT Press.

Jas, E. (2014). What Other Josquin? *Early Music History*, 33, 109–142.

Juola, P. (2015). The Rowling Case: A Proposed Standard Analytic Protocol for Authorship Questions. *Digital Scholarship in the Humanities*, 30 (suppl 1), pp. i100–i113.

van Kranenburg, P. (2006). Composer attribution by quantifying compositional strategies. In ISMIR Proceedings, pp. 375–376.

Levitt, M. (2001). Joyce, Yet Again. *Journal of Modern Literature*, 24(3), 507–511.

Love, H. (2002). *Attributing Authorship*. Cambridge: Cambridge University Press.

Mosteller, F., & Wallace, D. (1964). *Inference and Disputed Authorship: The Federalist*. Addison-Wesley.

Osthoff, H. (1962). Josquin des Prez, in MGG1, vol. 7. pp. 190–214.

Parncutt, R., & McPherson, G. E. (Eds.) (2002). *The science and psychology of music performance: Creative strategies for teaching and learning*. New York: Oxford University Press.

Patel, A. D., & Daniele, J. R. (2003). An empirical comparison of rhythm in language and music. *Cognition*, (87/1).

Rodin, J, & Sapp, Craig. (2010). The Josquin Research Project. http://josquin.stanford.edu, accessed on 5/5/2016.

Speiser, Jacqueline, & Gupta, Vishesh. "Composer Style Attribution". Project Report for CS 229: Machine Learning, Stanford University; 2013. cs229.stanford.edu/proj2013/CS229-FinalProject.pdf

Sparks, E. H. (1971). Problems of authenticity in Josquin's motets. Proceedings of the International Festival-Conference. edd. E.E. Lowinsky & B.J. Blackburn (London 1976), pp. 345–59.

Wegman, R. C. (2008). The Other Josquin. Tijdschrift van de Koninklijke Vereniging Voor Nederlandse Muziekgeschiedenis, 58(1), 33–68.

An Investigation of Multimodal Imagery Employed by Ear Players without Absolute Pitch

Anne-Marie Hildebrandt[1], Yo-Jung Han[2]
[1]Music Education, The Pennsylvania State University, University Park, PA, USA
[2]Music Education, The Pennsylvania State University, University Park, PA, USA

Playing by ear is considered a foundational musical skill, yet few studies examine the mental processes involved in it. As externalizing internal music involves multimodal imagery, this study aims to illuminate the mental imagery musicians without absolute pitch may experience while playing by ear. Employing a qualitative case study design, we will purposively select six proficient ear players without absolute pitch, enculturated in Western tonal musical traditions, representing multiple learning backgrounds and musical instruments.

The first two phases of data collection will facilitate participants' metacognition, often a challenge with automatized processes such as playing by ear. First, participants will be asked to play several melodic passages by ear, designed to require varying levels of abstraction, and verbalize their mental imagery using a think-aloud protocol. Participants will then refine their descriptions for one week in a reflective journal, supplemented by daily prompts from researchers. Finally, semi-structured interviews will allow researchers to gather information about participants' backgrounds and to clarify their previous descriptions. Participants will then check the accuracy of researchers' transcription and interpretation of their descriptions. While analyzing data using systematic coding techniques, we will be particularly cognizant of participants' usage of imagery as they employ strategies which may include the tracing of contour through intervals, the use of intervals to relate pitches to the tonic, a sense of tonal hierarchy, and acquired motor associations. Based on existing literature, we expect formal and informal learning experiences, exposure to music notation, and features of participants' musical instruments to influence the multimodal imagery used while playing by ear.

Towards the Objective Assessment of Music Performances

Chih-Wei Wu, Siddharth Gururani, Christopher Laguna, Ashis Pati, Amruta Vidwans, Alexander Lerch

Center for Music Technology, Georgia Institute of Technology, USA
{cwu307, siddgururani, claguna3, ashis.pati, avidwans, alexander.lerch}@gatech.edu

ABSTRACT

The qualitative assessment of music performances is a task that is influenced by technical correctness, deviations from established performance standards, and aesthetic judgment. Despite its inherently subjective nature, a quantitative overall assessment is often desired, as exemplified by US all-state auditions or other competitions. A model that automatically generates assessments from the audio data would allow for objective assessments and enable musically intelligent computer-assisted practice sessions for students learning an instrument. While existing systems are already able to provide similar basic functionality, they rely on the musical score as prior knowledge. In this paper, we present a score-independent system for assessing student instrument performances based on audio recordings. This system aims to characterize the performance with both well-established and custom-designed audio features, model expert assessments of student performances, and predict the assessment of unknown audio recordings. The results imply the viability of modeling human assessment with score-independent audio features. Results could lead towards more general software music tutoring systems that do not require score information for the assessment of student music performances.

I. INTRODUCTION

Music performance, according to Palmer, is one of the most complex serial actions produced by human beings, requiring the interpretation of musical ideas, the planning of the retrieved musical units, and the transformation of these thoughts into movements (Palmer, 1997). The qualitative assessment of performances by teachers and peers is, despite its inherent difficulty due to the subjective nature of the task, an essential part of music education. The teacher has to provide structured quality assessment, possibly quantifying different aspects of a performance and thus providing systematic feedback in order to facilitate improvement and reach the learning outcomes. However, as pointed out by Thompson and Williamon (Thompson & Williamon, 2003), the bias of the evaluators and the highly correlated categories in the structured assessment could impact the discriminability and the fairness of this approach. A computational approach that models the human cognition of the music performances and provides consistent and reproducible feedback might be a potential solution to this issue. It might also be used as a tool to provide feedback to students during practice sessions without an instructor.

Approaches and tools from the research field Music Information Retrieval (MIR) are being utilized more and more in software solutions for music education (Dittmar, Cano, Abeßer, & Grollmisch, 2012). With the advancement in research topics such as source separation (Huang, Kim, Hasegawa-Johnson, & Smaragdis, 2014) and music transcription (Benetos, Dixon, Giannoulis, Kirchhoff, & Klapuri, 2013), different music learning systems with reliable functionalities can be created, offering objective and repeatable evaluation to the users. Commercial software such as SmartMusic (http://www.smartmusic.com Last access: 2016/04/24) and Yousician (https://get.yousician.com Last access: 2016/04/24) are examples of such systems.

Since most of these systems require the score as additional input, their applications are limited and frequently depend on proprietary curated content provided by the manufacturer. In this paper, we explore the idea of building a score-independent system. A set of well-established audio features used in MIR tasks (Tzanetakis & Cook, 2002) is compared to a set of custom-designed features in a machine learning based regression model predicting the assessments of human experts. The goal is to investigate whether score-independent descriptors can be meaningful for the general assessment of student music performances. This paper is structured as follows: Sect. 2 introduces the related work on music performance assessment. The details of the dataset used in this work are described in Sect. 3. The methodology and the experimental setups are mentioned in Sect. 4. Finally, the experiment results and conclusions are presented in Sect. 5 and 6, respectively.

II. RELATED WORK

Music performance analysis is a research field that involves the observation, extraction, and modeling of important parameters in music performances. Early research focused on the analysis of symbolic data collected from external sensors or MIDI devices. More recently, the focus has gradually shifted to the analysis of audio recordings. For example, Abeßer et al. proposed a system that automatically assesses the quality of vocal and instrumental performances of 9^{th} and 10^{th} graders (Abeßer, Hasselhorn, Dittmar, Lehmann, & Grollmisch, 2013). The assessment is obtained by modeling the relationship between score-based features and the experts' ratings. The rating ranges from 1 to 4: 1 being the best, and 4 being the worst performance quality. A four-class classifier is trained to assess the performance, and the evaluation results show that the system is able to classify the performances although exhibiting some confusion between adjacent ratings. Another example is the score-informed piano tutoring system presented by Fukuda et al. (Fukuda, Ikemiya, Itoyama, & Yoshii, 2015), which applies automatic music transcription and audio to score alignment to detect the mistakes in the user's performance. The system also includes a score-simplification functionality to motivate the users by reducing the difficulty of a given score. The evaluation results show that the system can transcribe the audio input with high accuracy, and highlight the mismatches between the score and the performance with occasional octave errors.

In the studies mentioned above, a score is usually a prerequisite for the automatic assessment. However, in certain use cases, such as free practice or improvisation, these systems are not directly applicable since no score is available.

Nakano presented an automatic system that evaluates the singing skill of the users (Nakano, Goto, & Hiraga, 2006). The system is trained based on the extracted pitch interval accuracy and vibrato features without any score input. The evaluation results show that the system is able to classify the performance into two classes (good or poor) with high accuracy. Mion and De Poli proposed a system that classifies music expressions based on score-independent audio features (Mion & De Poli, 2008). With instantaneous and event-based features such as spectral centroid, residual energy, and notes per second, the system is able to recognize four different musical expressions of violin, flute, and guitar performances. These examples show the potential of analyzing a recorded music performance without the need for the underlying score.

III. DATASET

The dataset used for this study is kindly provided by the Florida Bandmasters Association (FBA). The dataset contains 3344 audio recordings of the 2013–2014 Florida all-state auditions with accompanying expert assessments. The participating students are divided into three groups, namely middle school (7th and 8th grade), concert band (9th and 10th grade), and symphonic band (11th and 12th grade). A total number of 19 types of instruments are played during the auditions. More details are shown in Table 1. All of the auditions are recorded at a sampling rate of 44100 Hz and are encoded with MPEG-1 Layer 3.

For pitched instruments, each audition session includes 5 different exercises, which are lyrical etude, technical etude, chromatic scale, 12 major scales, and sight-reading. For percussion instruments, the audition session also includes 5 different exercises, which are mallet etude, snare etude, chromatic scale (xylophone), 12 major scales (xylophone), and sight-reading (snare). Each exercise is graded by human experts with respect to different assessment categories, such as musicality, note accuracy, rhythmic accuracy, tone quality, artistry, and articulation, etc. In our experiments, all of the ratings are normalized to a range between 0 and 1, with 0 being the minimum and 1 being the maximum score.

To narrow the scope of this study, only a small subset of this original dataset is used. This subset includes the recordings of one exercise from one pitched instrument (alto sax) and one percussion instrument (snare drum) performed by middle school students. These two instruments have been selected because they have relatively higher number of recordings in the dataset (122 and 98, respectively).

Table 1. Statistics of the 2013-2014 FBA audition dataset

	# Files	Total Duration (mins)	Avg. Duration (mins)
Middle School	1099	5460	4.96
Concert Band	1046	5280	5.04
Symphonic Band	1199	6540	5.45
List of All Instruments			
Alto Sax, Baritone Sax, Bass Clarinet, Bass Trombone, Bassoon, Bb Clarinet, Bb Contrabass Clarinet, Eb Clarinet, English Horn, Euphonium, Flute, French Horn, Oboe, Percussion, Piccolo, Tenor Sax, Trombone, Trumpet, Tuba			

IV. EXPERIMENTS

A. Feature Extraction

To represent the audio signals in the feature space, two types of features are extracted: 1) baseline features and 2) designed features.

The baseline features (d = 17, d is the dimensionality) are computed block-by-block with a window size of 1024 samples and a hop size of 256 samples. A Hann window is applied to each block. The baseline features include: spectral centroid, spectral rolloff, spectral flux, zero-crossing rate, and 13 Mel Frequency Cepstral Coefficients (MFCCs). The features are implemented according to the definitions in (Lerch, 2012). To represent each recording with one feature vector, a two-stage feature aggregation process has been applied. In the first stage, the block-wise features within a 250ms texture window are first aggregated by their mean and standard deviation. In the second stage, all of these meta-features are aggregated again into one single vector with their mean and standard deviation, resulting in one baseline feature vector (d = 68) per recording.

The designed features for the percussion instruments are used to capture the rhythmic aspects of the audio signal. The resulting features (d = 18) per recording are shown as follows:

- Inter-Onset-Interval (IOI) histogram statistics (d = 7): This set of features is designed to describe the rhythmic characteristics of the played onsets. An IOI histogram is computed with 50 bins. Next, the standard statistical measures crest, skewness, kurtosis, rolloff, flatness, tonal power ratio, and the histogram resolution are extracted.
- Amplitude histogram statistics (d = 11): This set of features is designed to capture the amplitude variations of the played onsets. The amplitude of the waveform is first converted into dB, and the amplitude histogram is calculated with 50 bins. Next, the same statistical measures as above are extracted. Additionally, the length of the exercise, the standard deviation of the amplitude in both linear and dB scale, and the Root Mean Square (RMS) of the entire waveform are included as features.

The designed features for the pitched instruments are intended to describe various dimensions of the performances, namely the pitch, dynamics, and timing characteristics. Most of these features are extracted at the note-level after a simple segmentation process based on the quantized pitch contour. The pitches are detected using an autocorrelation function based pitch tracker. The note-level features are:

- Note steadiness (d = 2): These two features are designed to find fluctuations in the pitch of a note. For each note, the standard deviation of pitch values and the percentage of values deviating more than one standard deviation from the mean are computed.

- Amplitude deviation (d = 1): This feature aims to find the uniformity of the amplitude of a note. For each note, the standard deviation of the RMS is computed.
- Amplitude envelope spikes (d = 1): This feature describes the spikiness of the note amplitude over time. The number of local maxima of the smoothed derivative of the RMS is computed per note.

Once the above features are extracted, their mean, maximum, minimum, and standard deviation across all the notes are calculated to represent the recording. In addition, the following exercise-level features are extracted:

- Average pitch accuracy (d = 1): This feature shows the consistency of the notes played. The histogram of the pitch deviation from the closest equally tempered pitch is extracted with a 10 cent resolution. The area under the window (width: 30 cent) centered around the highest peak is considered as the feature.
- Percentage of correct notes (d = 1): For this feature, each note is labeled either correct or incorrect, and the percentage of correct notes across the entire exercise is computed as the feature. A note is labeled correct if the percentage of pitch values with a deviation from the mean pitch is lower than a pre-defined threshold.
- Timing accuracy (d = 7): These features are computed from the IOI histogram of the note onsets. Note onsets are computed from the pitch contour. The features used are the same as the features that were used for percussive instruments.

The resulting feature vector for the designed features for pitched instruments has the dimension d = 25.

B. Feature Extraction

Using the extracted features from the audio signals, a Support Vector Regression (SVR) model with a linear kernel function is trained to predict the human expert ratings. The libsvm (Chang & Lin, 2011) implementation of this model is used with default parameter settings. A Leave One Out cross-validation scheme is adopted to train and evaluate the models.

C. Experimental Setup

To compare the effectiveness of the baseline features versus the designed features, two sets of experiments are conducted. In the first set of experiments, the baseline and designed features are extracted from the pitched instrument recordings and used to build the regression models for predicting labels such as artistry, musicality, note accuracy, rhythmic accuracy, and tone quality. In the second set of experiments, the same procedure is used to predict musicality, note accuracy, and rhythmic accuracy for the percussion instrument recordings. An outlier removal process is included for all the experiments. This process removes the training data with the highest prediction residual (prediction minus actual rating) and is repeated until 5% of the dataset is eliminated. By removing the outliers, the regression models should be able to better capture the underlying patterns in the data.

D. Evaluation Metrics

The performances of the models are evaluated by the following standard statistical metrics: the correlation coefficient r and its corresponding p-value, the R^2 value, and the standard error. These metrics are typically used to measure the strength of the regression relationship between the predictions and the actual ratings. More details of the mathematical formulations can be found in (McClave & Sincich, 2003).

V. RESULTS AND DISCUSSION

The results of the first and second set of experiments are shown in Table 2. Most of the entries have $p \ll 0.05$ except for artistry in the baseline features and note accuracy in both features of the pitched instrument; therefore, the p-values are not shown in the table.

The following trends can be observed: first, the baseline feature set is not able to capture enough meaningful information to create a usable model.

Second, compared to the baseline features, the designed features lead to general improvements in all the metrics for both pitched and percussion instruments. In most cases, the R^2 values increase by at least 30%, illustrating the effectiveness of the designed features at characterizing the music performances.

Third, musicality is the assessment that is modeled best for both pitched and percussion instruments. The highest correlation coefficient and R^2, 0.7307 and 0.5254 respectively, are achieved when using designed features to predict musicality for pitched instrument recordings. This could be explained by the fact that musicality is a relatively abstract description that covers most aspects of a performance, and therefore, it is likely to be related to the general quality of the performance, making it relatively easy to model. Table 3 shows the result of an inter-label investigation, correlating one type of rating with the others. It is found that musicality tends to be highly correlated with the other assessment categories. This result further confirms the relationship between musicality and overall performance quality.

Fourth, note accuracy for the pitched instrument is the worst performing category. Compared to a percussion instrument, more information might be required to model the note accuracy of a pitched instrument properly. The proposed features might be unsuitable for capturing the relevant information. This result could imply the necessity of including score information in order to improve the model of certain categories.

Last but not least, the baseline features did not perform better than the designed features in predicting the tone quality. This seems to contradict the intuition that timbre features are directly related to the tone quality. On the one hand, this could mean that other, not extracted low-level characteristics such as the initial attack phase of each note might have larger impact on the tone quality than the spectral envelope modeled by our features. On the other hand, this result might suggest a connection between the perceived tone quality and the high-level information captured by the designed features, such as note steadiness and pitch accuracy.

Table 2. Experiment results of the regression models

Inst	Feature	Label	r	R^2	Std. Err
Pitched (Alto Sax)	Baseline	Artistry	0.1559	-0.6750	0.2025
		Musicality	0.4698	-0.0693	0.0914
		Note Acc.	0.1134	-0.5735	0.1765
		Rhyt. Acc.	0.4099	-0.0783	0.1674
		Tone Qual.	0.3659	-0.2372	0.1056
	Designed	Artistry	0.4548	0.1635	0.1385
		Musicality	0.7307	0.5254	0.0627
		Note Acc.	0.1578	-0.0840	0.1465
		Rhyt. Acc.	0.6252	0.3727	0.1195
		Tone Qual.	0.5249	0.2350	0.0810
Percussion (Snare)	Baseline	Musicality	0.4537	0.0661	0.1709
		Note Acc.	0.4853	0.0695	0.1592
		Rhyt. Acc.	0.3835	-0.0025	0.1685
	Designed	Musicality	0.6489	0.4174	0.1293
		Note Acc.	0.5674	0.2393	0.1384
		Rhyt. Acc.	0.5467	0.2745	0.1497

VI. CONCLUSION

In this paper, a system that automatically assesses student music performances based on score-independent audio features is presented. With the presented features the system is, to a certain degree, able to model and predict the ratings given by human experts without prior knowledge of the underlying scores. The results of the experiments show that, for both pitched and percussion instruments, the designed features perform better than the baseline features. Overall, the results imply the general feasibility of using score-independent features in assessing student music performances, although the presented features only capture part of the performance characteristics to be modeled.

The challenges and future directions of this research are: first, the current dataset, in spite of being diverse, only contains a few samples for specific combinations of instrument and group. To build a generic model, more data is needed. Second, although the designed features are shown to be useful, they require a lot of domain knowledge and might not be directly applicable to other datasets. An alternative solution is to apply a feature learning method such as Sparse Coding (Abdallah & Plumbley, 2006), which could automatically learn relevant features and potentially achieve a higher performance. Finally, to further investigate the necessity of the musical score especially in certain assessment categories such as note accuracy, score-based features should be explored and compared in the future.

ACKNOWLEDGMENT

The authors would like to thank the Florida Bandmasters Association for providing the dataset used in this study.

Table 3. Inter-label cross-correlation results for musicality

	Artistry	Musicality	Note	Rhythmic	Tone
Pitched Instrument					
Musicality	0.7352	1	0.6439	0.7967	0.7037
Percussion Instrument					
Musicality	N/A	1	0.6790	0.8090	N/A

REFERENCES

Abdallah, S. A., & Plumbley, M. D. (2006). Unsupervised analysis of polyphonic music using sparse coding. *IEEE Transactions on Neural Networks*, *17*(1), 179–196.

Abeßer, J., Hasselhorn, J., Dittmar, C., Lehmann, A., & Grollmisch, S. (2013). Automatic quality assessment of vocal and instrumental performances of ninth-grade and tenth-grade pupils. In *Proceedings of the International Symposium on Computer Music Multidisciplinary Research (CMMR)* (pp. 975–988).

Benetos, E., Dixon, S., Giannoulis, D., Kirchhoff, H., & Klapuri, A. (2013). Automatic music transcription: challenges and future directions. *Journal of Intelligent Information Systems*.

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM : a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, *2*(3), 27:1–27:27.

Dittmar, C., Cano, E., Abeßer, J., & Grollmisch, S. (2012). Music Information Retrieval Meets Music Education. In *Multimodal Music Processing* (Vol. 3, pp. 95–120). Schloss Dagstuhl--Leibniz-Zentrum fuer Informatik.

Fukuda, T., Ikemiya, Y., Itoyama, K., & Yoshii, K. (2015). A Score-informed Piano Tutuoring System with Mistake Detection and Score Simplification. In *Proceedings of Sound and Music Computing Conference (SMC)*.

Huang, P.-S., Kim, M., Hasegawa-Johnson, M., & Smaragdis, P. (2014). Singing-voice separation from monaural recordings using deep recurrent neural networks. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*.

Lerch, A. (2012). *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics*. John Wiley & Sons.

McClave, J. T., & Sincich, T. (2003). *Statistics* (9th ed.). Upper Saddle River, NJ: Prentice Hall.

Mion, L., & De Poli, G. (2008). Score-Independent Audio Features for Description of Music Expression. *IEEE Transactions on Audio, Speech, and Language Processing*, *16*(2), 458–466.

Nakano, T., Goto, M., & Hiraga, Y. (2006). An automatic singing skill evaluation method for unknown melodies using pitch interval accuracy and vibrato features. In *Proceedings of International Conference on Spoken Language Processing (ICSLP)* (pp. 1706–1709).

Palmer, C. (1997). Music performance. *Annual Review of Psychology*, *48*, 115–138. http://doi.org/10.1146/annurev.psych.48.1.115

Repp, B. H. (1996). Patterns of note onset asynchronies in expressive piano performance. *The Journal of the Acoustical Society of America*, *100*(6), 3917–3932.

Shaffer, L. (1984). Timing in solo and duet piano performances. *The Quarterly Journal of Experimental Psychology*, *36*(4), 577–595.

Thompson, S., & Williamon, A. (2003). Evaluating evaluation: Musical performance assessment as a research tool. *Music Perception*, *21*(1), 21–41.

Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on Audio, Speech and Language Processing*, *10*(5), 293–302.

Predicting Missing Music Components with Bidirectional Long Short-Term Memory Neural Networks

I-Ting Liu and Richard Randall

Abstract**—**Successfully predicting missing components from complex multipart musical textures has attracted researchers of music information retrieval and music theory. Solutions have been limited to either two-part melody and accompaniment (MA) textures *or* four-part Soprano-Alto-Tenor-Bass (SATB) textures. This paper proposes a robust framework applicable to both textures using a Bidirectional Long-Short Term Memory (BLSTM) recurrent neural network. The BLSTM system was evaluated using frame-wise accuracies on the Nottingham Folk Song dataset and J. S. Bach Chorales. Experimental results demonstrated that BLSTM significantly outperforms other neural-network based methods by 4.6% on average for four-part SATB and two-part MA textures. The high accuracies obtained with BLSTM on both textures indicated that it is the most robust and applicable structure for predicting missing components from multi-part musical textures.

Keywords**—**bidirectional long-short term memory neural network, multipart musical textures, music information retrieval

I. Introduction

THIS paper presents a method for predicting missing components from complex multipart musical textures. Specifically, we examine two-part melody and accompaniment (MA) and Soprano-Alto-Tenor-Bass (SATB) chorale textures. We treat each voice as a part (*e.g.* the melody of the MA texture or the Soprano of the SATB texture) and the problem we address is given an incomplete texture, how successfully can we generate the missing part. This project proposes a robust approach that is capable of handling both textures elegantly and has applications to any style of music. Predictions are made using a Bidirectional Long-Short Term Memory (BLSTM) recurrent neural network that is able to learn the relationship between components, and can thus be trained to predict missing components. This work demonstrates the capability of the BLSTM system by conducting experiments on the two tasks mentioned above with two distinct datasets.

Analyzing music with the aid of computer programs has attracted researchers of music information retrieval and music theory over the past twenty years. Music (especially western tonal music) has always been regarded as a kind of art with rigorous formalization [1]. Various complex rules regulate how notes can be and cannot be played together in complex multipart textures. Such rules change over time and are subject to multiple factors [2]. As artificial intelligence and machine-learning research advances, it is natural that computer scientists apply such technique to music analysis in order to elucidate these rules [3]. Two popular tasks investigated in this area are (1) generating chord accompaniments for a given melody in a two-part MA texture and (2) generating a missing voice for an incomplete four-part SATB texture. Successfully accomplishing either task manually is time-consuming and requires considerable style-specific knowledge and the applications discussed below are designed to automate and help non-professional musicians compose and analyze music.

Approaches that treat these problems can be categorized into two types according to the level of human engagement in discovering and applying music rules. Early works that handle incomplete four-part SATB textures were mostly knowledge-based models. Steels [4] proposed a representation system to encode musical information and exploit heuristic search, which takes the form of if-then musical rules that specify solutions under different conditions to generate voices. Ebcioglu built CHORAL, a knowledge-based system that includes over 350 rules modeling the style of Johann Sebastian Bach [5]. Due to the large number of rules involved, some studies modeled the problem as a constraint satisfaction problem, as was used by Pachet and Roy [6] on four-part textures and Ramirez et al. [7] on two-part textures. Knowledge-based genetic algorithms were also used as an alternative method to represent the rules. McIntyre [8] implemented a system that harmonizes user-defined melody in Baroque style, and Hall [9] presented a system that selects combinations of attributes to model the harmonization of J. S. Bach's chorales. Freitas and Guimaraes also implemented a system based on genetic algorithms in [10]. The fitness function and genetic operators rely on "music knowledges" to suggest chord progressions for given melodies.

While rules in knowledge-based systems have to be manually encoded into these systems, rules in probabilistic models and neural networks can be derived by training corpora without human intervention by the models. Hidden Markov Models (HMM) are one of the most common probabilistic models for the task of generating a chord sequence given melodies for two-part textures [11][12]. In HMM, a pre-selected dataset is used to train a transition probability matrix, which represents the probability of changing from one chord to another, and a melody observation matrix, the probability of encountering

R. Randall is with the School of Music and the Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, PA 15213 USA (412-268-4785; email: randall@cmu.edu).

I-T. Liu is with School of Music, Carnegie Mellon University, Pittsburgh, PA 15213 USA. (e-mail: itingl@andrew.cmu.edu).

each note when different chords are being played. The optimal chord sequence is then generated using dynamic programming, or Viterbi Algorithm. HMM are also used by Allan [13] [14] to harmonize four-part chorales in the style of J. S. Bach. In addition to HMM, Markov Model and Bayesian Networks are alternative models used for four-part textures by Biyikoglu [15] and Suzuki, et al. [16]. Raczynski, et al. [17] proposed a statistical model that combines multiple simple sub-models. Each sub-model captures different music aspects such as metric and pitch information, and all of them are then interpolated into a single model. Paiement, et al. [18] proposed a multi-level graphical model, which is proved to capture the long-term dependency among chord progression better than traditional HMM. One drawback of probabilistic models is that they cannot correctly handle data that are not seen in training data. Chuan and Chew [19] reduced this problem by using a hybrid system for style-specific chord sequence generation with statistical learning approach and music theory. In [20], Chuan compared and evaluated rule-based Harmonic Analyzer [21], probabilistic-model based MySong [18], and the hybrid system proposed in [19] on the task of style-specific chord generation for melodies.

Neural networks have also been used. Gang, et al. [22] were one of the earliest that used neural networks to produce chord harmonization for given melodies. Sequential neural network consisted of a sub-net that learned to identify chord notes for the melody in each measure, and the result was fed into the network to learn the relationship between melodies and chords. The network was later adopted in real-time application [23][24]. Consisting of 3 layers, the input layer takes pitch, metric information, and the current chord context, and the output layer predicts the next chord. Cunha, et al. [25] also proposed a real-time chord harmonization system using multi-layer perceptron (MLP) neural networks and a rule-based sequence tracker that analyzes the structure of the song in real-time, which provides additional information on the context of the notes being played.

Hoover, et al. [26] used two Artificial Neural Networks (ANN) to model the relationship between melodies and accompaniment as a function of time. The system was later extended to generate multi-voice accompaniment by increasing the size of the output layer in [27]. Bellgard and Tsand [3] trained an effective Bolzmann machine and incorporated external constraints so that harmonization follows the rules of chorales. Fuelner developed a feed-forward neural network that harmonizes melodies in specific styles in [28]. De Prisco, et al. [29] proposed a neural network that finds appropriate chords to harmonize given bass lines in four-part SATB chorales by combining three base networks, each of which models contexts of different time lengths.

Although these previous studies provide valuable insights, a number of constraints exist in their applications. Most rules encoded in knowledge-based systems are style-specific, making them hard to apply to other types of music efficiently. Probabilistic models and neural networks, on the other hand, provide a much more adaptable solution that can be applied to music of different styles by learning rules from different styles of training data. Nevertheless, many of the probabilistic models can only handle music pieces of fixed length. In addition, the transition matrix of probabilistic models has to be learned using specific music representation (*e.g.* chords) and cannot be generalized to other representations (*e.g.* SATB). Moreover, probabilistic models tend to ignore long-term dependency among music components as they mainly focus on local transitions between two consecutive components. Existing studies using neural networks captured long-term dependencies in music and also are capable of dealing with music pieces of arbitrary lengths. However, neural networks have been notoriously hard to train, and their ability to utilize long-term information was limited until the introduction of Long-Short Term Memory (LSTM) cells.

The current study builds on the LSTM model and has the advantages of several other models without their restrictions. Like probabilistic models, neural networks can learn music "rules" of different styles of music from training data. Yet unlike probabilistic models where a transition matrix among chords is required, neural networks enable the model to deal with flexible polyphonic accompaniment that does not necessarily have to be in the form of chords. Such structure also allows the model to incorporate additional musical quantities easily by adjusting the number of input neurons. Finally, adding *bidirectional links* and LSTM cells improves a neural network's ability to employ additional timing information. All of the above contributes to the fact that the proposed BLSTM model is flexible and effective in generating the missing component in an incomplete texture.

II. Background

A. Feed-Forward Neural Network

The most common neural network is a feed-forward multi-layered perceptron (MLP) network as shown in Fig. 1, which consists of three layers: an input layer, a hidden layer, and an output layer. The network is usually trained via back-propagation. Feed-forward neural networks were used extensively in music-related research such as harmony generation [22][25], onset detection [30], and algorithmic composition [31].

Though being the simplest among various neural network structures, feed-forward neural networks could not effectively capture rhythmic patterns and music structures in music owing to the fact that they do not have a mechanism to keep track of the notes played in the past. Since each input frame is processed individually, feed-forward neural networks are totally deterministic unless the context is provided to the network such as using a sliding window as in [30].

B. Recurrent Neural Network

Another way to present past information is to add recurrent links to the network, resulting in a recurrent neural network (RNN). A RNN has at least one feedback connection from one or more of its units to another unit, forming cyclic paths in the network. Fig. 2 shows a simple example of a RNN with an input layer, a hidden layer, and a output layer with recurrent links from the output layer to the input layer. RNNs are known to be able to approximate dynamical systems due to internal states that act as internal memory to process sequence of inputs through time. The fact that music is a complex structure that has both short-term and long-term dependency just as language models makes RNN an ideal structure for solving music-related problems. Mozer [32] used a fully-connected RNN to generate music note-by-note. Boulanger, Lewandowski, *et al*. also developed an RNN-based model to recognize chords in audio files [33] and construct polyphonic music [34] by using restricted Boltzmann machine (RBM) and recurrent temporal RBM (RTRBM).

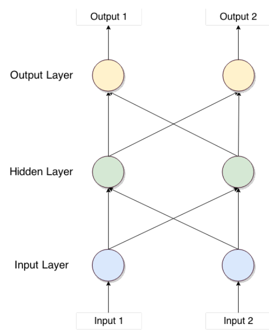


Fig. 1. A Multi-layer Feed-forward Neural Network

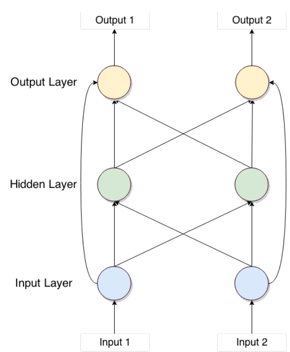


Fig. 2 A Multi-layer Recurrent Neural Network

C. Bidirectional Recurrent Neural Network

Standard RNNs process inputs in temporal order, and their outputs are mainly based on previous context. One way to include future information in the network is to use a bidirectional recurrent neural network (BRNN). Fig. 3 shows the structure of a BRNN with two hidden layers (one forward states, one backward states) unfolded in time. In a BRNN, two separate hidden layers are used, both connected to the same inputs and outputs. One of the layers processes inputs forward in time, while the other one processes inputs backward. Therefore, for each point of a given sequence, it could access complete temporal information before and after. BRNN has been used successfully to classify speech data in [35]. Eyben, *et al.* [36] also achieved impressive results on onset detection using BRNN. For the complete algorithm for training BRNN, please refer to [37].

D. Long Short Term Memory (LSTM)

Although BRNNs have access to both past and future information, they have been notoriously hard to train because of "vanishing gradients," [38] a problem commonly seen in RNNs when training with gradient based methods. Gradient methods, such as Back-Propagation Through Time (BPTT) [39], Real-Time Recurrent Learning (RTRL) [40] and their combinations, update the network by flowing errors "back in time." As the error propagates from layer to layer, it tends to either explode or shrink exponentially depending on the magnitude of the weights. Therefore, the network fails to learn long-term dependency between inputs and outputs. Tasks with time lags that are greater than 5-10 time steps are already difficult to learn, not to mention that dependency of music usually spans across tens to hundreds of notes in time, which contributes to music's unique phrase structures. Long short-term memory (LSTM) [38] algorithm was designed to tackle the error-flow problem.

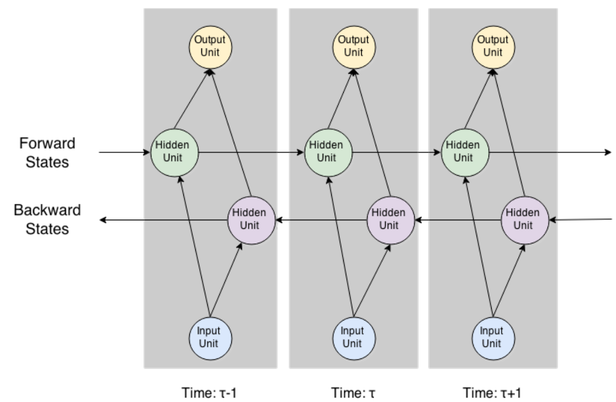


Fig. 3 A Bidirectional Recurrent Neural Network (BRNN) unfolded in time. Image from [37].

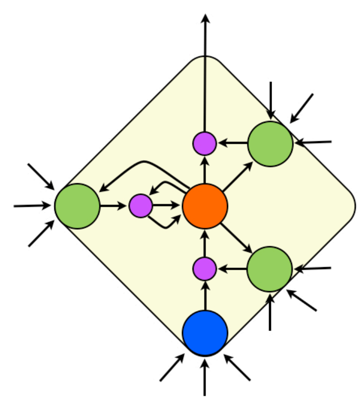


Fig. 4 A LSTM block that contains one linear cell (orange) and three non-linear gating units (green). Image from [41]

In a LSTM hidden layer, fully-connected memory blocks replace nonlinear units that are often used in feed-forward neural network. Fig. 4 shows an LSTM block. The core of a memory block is a linear cell (orange) that sums up the inputs, which has a self-recurrent connection of fixed weight 1.0, preserving all previous information and ensuring they would not vanish as they are propagated in time. A memory block also contains three sigmoid gating units: input gate, output gate, and forget gate. An input gate learns to control when inputs are allowed to pass into the cell in the memory block so that only relevant contents are remembered; an output gate learns to control when the cell's output should be passed out of the block, protecting other units from interference from current irrelevant memory contents; a forget gate learns to control when it is time to forget already remembered value, *i.e.* to reset the memory cell. When gates are closed, irrelevant information does not enter the cell and the state of the cell is not altered. The outputs of all memory blocks are fed back recurrently to all memory blocks to remember past values.

In this project, we use Bidirectional Recurrent Neural Networks as their recurrent links grant the network access to both information in the past and in the future. We also use LSTM cells in the network to avoid vanishing gradient problems, the details of which are covered below.

III. Methods

A. Music Representation

MIDI files are used as input in both training and testing phases in this project. Multiple input and output neurons are used to represent different pitches. At each time, the value of the neuron associated with the particular pitch played at that time is 1.0. The values of the rest of the neurons are 0.0. We avoid distributed encodings and other dimension reduction techniques and represent the data in this simple form because this representation is common and assumes that neural networks can learn a more distributed representation within hidden layers. Monophonic inputs are used because of their adaptability to both monophonic and polyphonic data. We leave further consideration and evaluation of distributed representation to future work.

The music is split into time frames and the length of each frame depends on the type of music. Finding missing music components can then be formulated as a supervised classification problem. For a song of length t_1, for every time t from t_0 to t_1, given input $x(t)$, the notes played at time t, find the output $y(t)$, which is the missing component we try to predict. In other words, for two-part MA textures, $y(t)$ is the chord played at time t, while for four-part SATB textures, $y(t)$ is the pitch of the missing part at time t.

B. Generating Accompaniment in Two-Part MA Texture

1) Input and Output

The MIDI files are split into eighth-note fractions. The inputs at time t, $x(t)$, are the notes of the melody played at time t. Instead of representing the notes by their MIDI number, which spans the whole range of 88 notes on a keyboard, we used pitch-class representation to encode pitches into their corresponding pitch-class number. Pitch class, also known as "chroma," is the set of all pitches regardless of their octaves. That is, all C notes (C*0*, C*1*. . . C_n, *etc.*) are all classified as pitch-class C. All notes are represented with one of the 12 numbers corresponding to the 12 semitones in an octave. In addition to pitch-class information, two additional values are added as inputs: Note-Begin unit and Beat-Onunit. In order to be able to tell when a note ends, a Note-Begin unit is used to differentiate two consecutive notes of the same pitch from one note that is held for two time frames as was done by [31]. If the note in the melody begins at time n, the value of the Note-Begin unit is 1.0; if the note is sounding, but is held from previous time n or is not played at all, the value of the unit is 0.0. The Beat-Onunit, on the other hand, provides metric information to the network. If the time t is on a beat, the value of the Beat-Onunit is 1.0, otherwise 0.0. If it is at rest, the values of all input neurons are 0.0. The time signature information is obtained via meta-data in MIDI files.

The outputs at time t, $y(t)$, is the chord played at time t. We limit chord selection to major, minor, diminished, suspended, and augmented triads as in [12], resulting in 52 chords in total. The output units represent these 52 chords in a manner similar to the input neurons: the value of the neuron corresponding to the chord played at that time has a value of 1.0, and the values of the rest of the neurons are all 0.0.

2) Training the Network

The input layer has 14 input neurons: 12 neurons for each pitch in the pitch-class, one neuron for note-begin and one for Beat-Onunit. The network consists of two hidden layers for both forward and backward states, resulting in four hidden layers in total. In every hidden layer are 20 LSTM blocks with one memory cell. The output layer uses the softmax activation function and cross entropy error function as in [35]. Softmax function is a standard function for multi-class classification that squashes a K-dimensional vector x in the range of (0,1), which takes the form

$$\sigma(x)_j = \frac{e^{x_j}}{\sum_{k=1}^{K} e^{x_j}}, \text{ for } j = 1, ..., K$$

The softmax function ensures that all the output neurons sum to one at every time step, and thus can be regarded as the probability of the output chord given the inputs at that time. Each music piece is presented to the network one at a time, frame-by-frame. The network is trained via standard gradient-descent Back-Prorogation. A split of data is used as the validation set for early stopping in order to avoid over-fitting of the training data. If there is no improvement on the validation set for 30 epochs, training is finished and the network setting with the lowest classification error on the validation set is used for testing.

3) Markov Model as Post-Processing

The network trained in III-B2 can then be used to predict the chord associated with each melody note by choosing the output neuron that has the highest activation at each time point. However, the predicted chord at each time is independent of the chord predicted in the previous and succeeding time. While there are forward and backward links in the hidden layers of the network, there is no recurrent connections from the final neuron output to the network. The chord might sound good with the melody, but the transition from one chord to another might not make sense at all. In fact, how one chord transitions from and to the other typically follows specific chord-progression rules depending on different music styles. A bi-gram Markov Model is thus added to learn the probability of transitioning from each chord to possible successors independent of the melody, which will be referred to as the transition matrix. The transition matrix is smoothed using linear interpolation with a unigram model. The model also learns the statistics of the start chords.

Instead of selecting the output neuron with the highest activations, the first k neurons with the highest activations are chosen as candidates. Dynamic programming is then used to determine the optimal chord sequence among the candidates using the previously learned transition matrix.

C. Generating the Missing Part in SATB Textures

1) Input and Output

We sample the melody at every eighth note for similar reasons as explained by [29]. Notes that are shorter in length are considered as passing notes and are ignored here. The inputs at time t, $x(t)$, are the pitches of the notes played at time t, spanning the whole range of 88 notes (A0, C8) on a keyboard, resulting in an 88-dimensional vector. If a note i is played at time t, the value of the neuron associated with the particular pitch is 1.0, *i.e*., $x_i(t)$=1.0. The number of non-zero elements in $x(t)$, which are the number notes played each time, ranges from one to three, depending on the number of voices present.

For the task of predicting the missing voice, either Soprano, Alto, Tenor or Bass, in a four-part texture where the other three voices are present, the input is three-part polyphonic music. In this case, there are at most three non-zero elements in x_t for every time t. If the task is to predict one missing voice given only one of the three other voices, there is at most one non-zero element in $x(t)$. Note that we do not add any additional information to the network about which voice is missing nor which voice(s) are given; the network induces such knowledge according to the input data and the target output data. The reason why we do not represent the notes with their pitch-class profile as we did when handling two-part MA texture is that the network depends on octave information to identify which voice the notes belong to. The outputs at time t, $y(t)$, is the predicted missing note at time t, which falls in the pitch range of any of the four voices, depending on the task specified by our training data. Similarly, the value of the neuron associated with the particular pitch played at the time t is 1.0, otherwise 0.0.

2) Training the Network

The network structure is the same as the one used in Section III-B2 except that the number of input neurons and output neurons are 88, and that we use 20 LSTM blocks for the first hidden layer and 50 LSTM blocks for the second hidden layer. Similar to what we did for two-part MA textures, each music piece is presented to the network one at a time, frame-by-frame. If the task is to generate one missing voice given any of the three other voices, then the three present voices are given to the network individually as if they are independent melodies. In this case, each music piece is actually presented to the network three times in total, and each time only one of the three voices is presented. Training is finished if there is no improvement on the validation data for 30 epochs.

3) Predict Missing Voice with the Trained Network

The trained network is ready to predict the missing voice by doing an 88-class classification on the input voice. At each time frame, the neuron with the highest activations is selected, and the pitch it represents is considered as the pitch of the missing voice.

IV. EVALUATION

A. Generating Missing Accompaniment in MA Texture

1) Dataset

The system's performance on two-part MA textures is evaluated using the Nottingham Dataset [42] transcribed from ABC format, which is also used in [34] for composing polyphonic music. The dataset consists of 1024 double-track MIDI files, with melody on one track and accompaniment on the other. The length of the pieces ranges from 10 seconds to 7.5 minutes, the median being 1 minute and 4 seconds. Those without accompaniment and those whose accompaniment are more complicated than simple chord progressions are discarded, resulting in 962 MIDI files comprising more than 1000 minutes, in total. Songs not in the key of C major nor A minor (874 of them) were transposed to C major/A minor after probabilistically determining their original key using Krumhansl-Schmuckler key-finding algorithm [43][44].

The chords were annotated at every beat or at every quarter note. Seventh chords were reduced to triads, and rests were replaced with previous chords. 60% of the dataset is selected randomly as training data, 20% as validation data, and 20% as testing data. Training finishes when validation accuracy does not improve for 30 epochs. All results for the training and testing sets were recorded at the time when the classification error on the validation set is lowest.

2) Effects of Including Metric Information in Input

Since the network learns the input melody as a sequence in time and has no access to information other than pitches, we added Beat-On flag to a frame when it is on a beat according to the time signature meta-data in MIDI files (Group iii and iv). We also added Note-Begin (Group ii and iv) to differentiate two consecutive notes of the same pitch from two distinctive notes, as mentioned in Section III-B1.

All three groups were sampled every eighth note, and the MIDI note range (50, 95) was used as the input range. Table I shows the classification accuracy of the three groups as well as the one where neither flag is provided as a reference. Two groups where Beat-On flag is added, Group iii and iv, perform significantly better than the groups without the beat information (Group i). This is consistent with the fact that chords always change on a beat or multiples of a beat. Therefore, such information is crucial to the timing of chord changes in the network. Note-Begin, on the other hand, does not seem to improve the accuracy, which is due to the fact that whether the note is held from the previous time or it is newly started does not affect chord choices.

TABLE I
CLASSIFICATION ACCURACY OF THE DATASET WHEN A NOTE-BEGIN FLAG, BEAT-ONFLAG, AND BOTH FLAGS ARE ADDED TO THE INPUTS.

	Training Set	Test Set
(i) Pitch Information only	72.88%	68.54
(ii) Note-Begin	72.11%	68.86
(iii) Beat-On	75.82%	70.34
(iv) Note-Begin and Beat-On	75.76%	**70.61**

TABLE II
CLASSIFICATION ACCURACY OF THE DATASET WHEN USING VARIOUS REPRESENTATIONS OF PITCHES AT VARIOUS SAMPLING RATES.

	Training Set	Test Set
(i) 8th Note + Melody Range	75.76%	70.65 %
(ii) 8th Note + Pitch Class	73.13%	**72.05 %**
(iii) 16th Note + Melody Range	73.10%	69.50 %
(iv) 16th Note + Pitch Class	74.02%	70.67 %

3) Choice of Data Representations

This experiment discusses how to represent data to the network to achieve the best performance. To see how different resolutions of the melody affects the chord prediction result, we evaluated the performance of the system using different frame lengths. "8th Note" or "16th Note" indicates the melodies and accompaniments were sampled every eighth note or sixteenth note. To see how melodies are represented to the network affects the performance, we represented the input to the network using only the actual pitch range that melody notes are played in, which is MIDI note 50 (D3) to 95 (B6) (Groups i and iii - "Melody Range"), and using pitch class representation (Groups ii and iv - "Pitch Class").

Since the network learns the input melody as a sequence in time and has no access to information other than pitches, we added Beat-On flags to a frame when it is on a beat according to the time signature in meta-data in MIDI files. We also added Note-Begin flags to differentiate two consecutive notes of the same pitch from two distinctive notes. Representing the melodies with their pitch-class number at every 8th note (Group ii) could correctly predict the missing chords approximately 72% of the time when both Note-Begin and Beat-On information are available, and significantly outperforms other representations. Table II shows the result.

4) Comparison with Other Approaches

We compared the architecture used in this paper with four other neural network architectures: Unidirectional LSTM, Bidirectional recurrent neural network (BRNN), Unidirectional recurrent neural network (RNN), and Multi-layer perceptron network (MLP). Neurons in BRNN, RNN and MLP networks were sigmoid neurons. The size of the hidden layers were selected so that the number of weights are approximately the same (around 32,000) for all of the networks as in [35]

Table III shows the classification accuracy and the number of epochs required to converge. All groups were sampled at every eighth note, and were provided with both metric information, (Note-On and Beat-On), during training and testing. Using approximately same number of weights, BLSTM performs significantly better than other neural networks and also converges the fastest.

TABLE III
CLASSIFICATION ACCURACY OF THE DATASET USING DIFFERENT NEURAL NETWORK ARCHITECTURES.

Network	Training Set	Test Set	Epochs
BLSTM	**75.76%**	**71.13 %**	103
LSTM	71.51%	67.57 %	130
BRNN	68.77%	68.86 %	136
RNN	68.33%	66.58 %	158
MLP	55.16%	54.66 %	120

B. Finding the Missing Part in Four-Part SATB Textures

1) Dataset

We evaluated our approach using 378 of J. S. Bach's four-part chorales acquired from [45]. MIDI files were all multi-tracked, one voice on each track. The average length of the pieces is approximate 45 seconds, the maximum and minimum being 6 minutes to 17 seconds, respectively. Among all chorales, 102 pieces are in minor mode. All of the chorales were transposed to C major/A minor using Krumhansl-Schmuckler key-finding algorithm [43]. As in section IV-A, 60% of the files were used as training set, 20% as test set, and 20% as validation set, resulting in 226, 76, 76 pieces respectively.

2) Predicting Missing Voice Given the Other Three Voices

Table IV shows the frame-wise classification accuracy of the predicted missing voices (Soprano, Alto, Tenor, or Bass) when the three other voices are given on training and test sets. The accuracy of predicting missing voices on the original non-transposed set is also listed for comparison. All songs were sampled at every eighth note. From the table, we can observe a few interesting phenomena. First, transposing the songs remarkably improves prediction accuracy in both training and test set. This is not surprising since transposing songs in advance reduces complexity. The same pre-processing is also used by [15] [34] [12]. Second, we see that the network could correctly predict Soprano, Alto, and Tenor approximately 70% of the time when the songs were transposed. Specifically, Alto seems to be the easiest to predict, while Bass is the most difficult.

3) Comparison with Other Approaches

Similar to our approach in Section IV-A4, the size of the hidden layers were selected so that the number of weights are approximately the same (around 63,000) for all of the networks. Table V shows the classification accuracy of the missing voices (either Soprano, Alto, Tenor, or Bass) when all of the three other voices are present.

TABLE IV
CLASSIFICATION ACURRASY OF THE PREDICTED MISSING VOICES, EITHER SOPRANO, ALTO, TENOR, OR BASS, WHEN THE THREE OTHER VOICES ARE GIVEN ON TRAINING AND TESTING SETS.

	Soparno		Alto	
	Training	Test	Training	Test
Not Transposed	69.15%	46.82%	63.61%	47.61%
Transposed	77.90%	71.52%	82.65%	**73.90%**

	Tenor		Bass	
	Training	Test	Training	Test
Not Transposed	47.25%	39.85%	45.40%	36.93%
Transposed	78.47%	69.76%	70.09%	61.22%

TABLE V
CLASSIFICATION ACCURACY OF THE PREDICTED MISSING VOICES WHEN THREE OTHER VOICES ARE GIVEN USING DIFFERENT NETWORK ARCHITECTURE.

	Soparno		Alto	
	Training	Test	Training	Test
BLSTM	84.88%	73.86 %	82.65%	73.90 %
BRNN	90.25%	**74.37%**	85.37%	**74.30 %**
LSTM	85.27%	70.39%	77.14%	70.45%
RNN	81.90%	72.29%	80.31%	71.73%
MLP	68.74%	66.54%	73.51%	70.03%
	Tenor		**Bass**	
	Training	Test	Training	Test
BLSTM	78.47%	69.76%	70.09%	61.22 %
BRNN	80.95%	**70.13%**	74.58%	**63.74%**
LSTM	73.84%	64.89%	65.86%	57.69%
RNN	75.48%	67.20%	69.68%	59.69%
MLP	68.85%	65.68%	58.58%	56.14%

From the result, we can see that BLSTM performs as well as BRNN on Soprano, Alto, and Tenor parts and significantly outperforms other neural-network based methods on all parts. It also shows that including future information by using bidirectional connection effectively improves accuracy by 3% on average no matter using LSTM cells (in BLSTM and LSTM) or logistic cells (in BRNN and RNN).

V. CONCLUSION

This paper has presented an approach to predicting missing music components for complex multipart musical textures using Bidirectional Long-Short Term Memory (BLSTM) neural networks. We demonstrated the flexibility and robustness of the system by applying the method to two distinctive but popular tasks in the computer-music field: generating chord accompaniment for given melodies in two-part MA textures and generating the missing voice in four-part SATB textures. The proposed approach is capable of handling music pieces of arbitrary length as well as various styles. In addition, the network could be used to generate missing music components of different forms, *i.e.* single notes for four-part SATB textures or chords for two-part MA textures, by simply altering the number of input and output neurons.

Two sets of experiments were conducted regarding the two tasks on two datasets of completely different styles, and issues that influence prediction accuracies were discussed. For the task of predicting chord accompaniment in two-part MA texture, the experimental results showed that the BLSTM network could correctly generate chords for given melodies 72% of the time, which is significantly higher than 68%, the best accuracy achieved by using other neural network based approaches.

We also discovered that representing the melodies using their pitch class profile yielded the best result. As for the problem of finding the missing voice in four-part SATB textures, our experiment demonstrated that a BLSTM network could correctly predict the missing voice approximately 70% of the time on average when three other voices are present. Putting the two experimental results together, BLTSM significantly outperforms all other neural-network based networks for two-part MA textures and performs as well as BRNN for four-part SATB textures showing that the BLSTM network is the optimal structure for predicting missing components from multi-part musical textures.

For future work, we will look into ways to improve the prediction accuracy. The transposition stage in this project can be improved by replacing Krumhansl-Schmuckler key-finding algorithm with other state-of-the-art methods. We may also look into alternatives to transposition such as encoding the inputs using distributed encodings or using intervals among the parts rather than their absolute pitches. In addition, all the system parameters are currently configured to maximize the results on the dataset used in this project. In the future, we will add pre-training to the network to find the optimal features based on music quantities, such as determining the optimal sampling rate according to the amount of activities in textures. For post-processing, we would like to retrain the network with various time delays to learn dependency among music components at different time lags. The capability of using Markov Models to refine results will also be explored by modeling other music quantities such as rhythm and transitions among individual notes. More data in different keys will also be gathered, and the Markov Model's ability of handling music in different keys will be further investigated by applying the Markov Model on each key individually. Finally, the proposed approach could be developed into an interactive system to aid song composition and arrangement. The capability of the network will be further investigated with music prediction tasks of other kinds of music and textures such as predicting the melody of one instrument in songs that involve multiple instruments.

REFERENCES

[1] F. Pachet and P. Roy, "Musical harmonization with constraints: A survey," *Constraints*, vol. 6, no. 1, pp. 7–19, 2001.
[2] H. V. Koops, J. P. Magalhaes, and W. B. De Haas, "A functional approach to automatic melody harmonisation," in *Proceedings of the first ACM SIGPLAN workshop on Functional art, music, modeling & design*. ACM, 2013, pp. 47–58.
[3] M. I. Bellgard and C.-P. Tsang, "Harmonizing music the Boltzmann way," *Connection Science*, vol. 6, no. 2-3, pp. 281–297, 1994.
[4] L. Steels, *Learning the craft of musical composition*. Ann Arbor, MI: MPublishing, University of Michigan Library, 1986.
[5] K. Ebciog˘lu, "An expert system for harmonizing four-part chorales," *Computer Music Journal*, pp. 43–51, 1988.
[6] F. Pachet and P. Roy, "Mixing constraints and objects: A case study in automatic harmonization," in *Proceedings of TOOLS Europe*, vol. 95. Citeseer, 1995, pp. 119–126.
[7] R. Ramirez and J. Peralta, "A constraint-based melody harmonizer," in *Proceedings of the Workshop on Constraints for Artistic Applications (ECAI98)*, 1998.
[8] R. A. McIntyre, "Bach in a box: The evolution of four part baroque harmony using the genetic algorithm," in *Evolutionary Computation, 1994. IEEE World Congress on Computational Intelligence., Proceedings of the First IEEE Conference on*. IEEE, 1994, pp. 852–857.
[9] M. A. Hall, "Selection of attributes for modeling Bach chorales by a genetic algorithm," in *Artificial Neural Networks and Expert Systems, 1995. Proceedings., Second New Zealand International Two-Stream Conference on*. IEEE, 1995, pp. 182–185.

[10] A. Freitas and F. Guimaraes, “Melody harmonization in evolutionary music using multiobjective genetic algorithms,” in *Proceedings of the Sound and Music Computing Conference*, 2011.
[11] H.-R. Lee and J.-S. Jang, “i-ring: A system for humming transcription and chord generation,” in *Multimedia and Expo, 2004. ICME’04. 2004 IEEE International Conference on*, vol. 2. IEEE, 2004, pp. 1031–1034.
[12] I. Simon, D. Morris, and S. Basu, “Mysong: automatic accompaniment generation for vocal melodies,” in *Proceedings of the SIGCHI Con- ference on Human Factors in Computing Systems*. ACM, 2008, pp. 725–734.
[13] M. Allan, “Harmonising chorales in the style of Johann Sebastian Bach,” *Master’s Thesis, School of Informatics, University of Edinburgh*, 2002.
[14] M. Allan and C. K. Williams, “Harmonising chorales by probabilistic inference,” *Advances in neural information processing systems*, vol. 17, pp. 25–32, 2005.
[15] K. M. Biyikoglu, “A markov model for chorale harmonization,” in *Preceedings of the 5 th Triennial ESCOM Conference*, 2003, pp. 81–84.
[16] S. Suzuki, T. Kitahara, and N. University, “Four-part harmonization using probabilistic models: Comparison of models with and without chord nodes,” *Stockholm, Sweden*, pp. 628–633, 2013.
[17] S. A. Raczyński, S. Fukayama, and E. Vincent, “Melody harmonization with interpolated probabilistic models,” *Journal of New Music Research*, vol. 42, no. 3, pp. 223–235, 2013.
[18] J.-F. Paiement, D. Eck, and S. Bengio, “Probabilistic melodic harmonization,” in *Advances in Artificial Intelligence*. Springer, 2006, pp. 218–229.
[19] C.-H. Chuan and E. Chew, “A hybrid system for automatic generation of style-specific accompaniment,” in *4th Intl Joint Workshop on Computational Creativity*, 2007.
[20] C.-H. Chuan, “A comparison of statistical and rule-based models for style-specific harmonization.” in *ISMIR*, 2011, pp. 221–226.
[21] D. Temperley and D. Sleator, “The temperley-sleator harmonic analyzer,” 1996.
[22] D. Gang and D. Lehmann, “An artificial neural net for harmonizing melodies,” *Proceedings of the International Computer Music Association*, 1995.
[23] D. Gang, D. Lehmann, and N. Wagner, “Harmonizing melodies in real-time: the connectionist approach,” in *Proceedings of the International Computer Music Association, Thessaloniki, Greece*, 1997.
[24] D. Gang, D. Lehman, and N. Wagner, “Tuning a neural network for harmonizing melodies in real-time,” in *Proceedings of the International Computer Music Conference, Ann Arbor, Michigan*, 1998.
[25] U. S. Cunha and G. Ramalho, “An intelligent hybrid model for chord prediction,” *Organised Sound*, vol. 4, no. 02, pp. 115–119, 1999.
[26] A. K. Hoover, P. A. Szerlip, and K. O. Stanley, “Generating musical accompaniment through functional scaffolding,” in *Proceedings of the Eighth Sound and Music Computing Conference (SMC 2011)*, 2011.
[27] A. K. Hoover, P. A. Szerlip, M. E. Norton, T. A. Brindle, Z. Merritt, and K. O. Stanley, “Generating a complete multipart musical composition from a single monophonic melody with functional scaffolding,” in *International Conference on Computational Creativity*, 2012, p. 111.
[28] J. Feulner, “Neural networks that learn and reproduce various styles of harmonization,” in *Proceedings of the International Computer Music Conference*. International Computer Music Association, 1993, pp. 236–236.
[29] R. DePrisco, A. Eletto, A. Torre, and R. Zaccagnino, “A neural network for bass functional harmonization,” in *Applications of Evolutionary Computation*. Springer, 2010, pp. 351–360.
[30] H. Goksu, P. Pigg, and V. Dixit, “Music composition using genetic algorithms (ga) and multilayer perceptrons (mlp),” in *Advances in Natural Computation*. Springer, 2005, pp. 1242–1250.
[31] P. M. Todd, “A connectionist approach to algorithmic composition,” *Computer Music Journal*, pp. 27–43, 1989.
[32] M. C. Mozer, “Neural network music composition by prediction: Exploring the benefits of psychoacoustic constraints and multi-scale processing,” *Connection Science*, vol. 6, no. 2-3, pp. 247–280, 1994.
[33] N. Boulanger-Lewandowski, Y. Bengio, and P. Vincent, “Audio chord recognition with recurrent neural networks.” in *ISMIR*, 2013, pp. 335–340.
[34] ——,“Modeling temporal dependencies in high-dimensional sequences: Application to polyphonic music generation and transcription,” *arXiv preprint arXiv:1206.6392*, 2012.
[35] A. Graves and J. Schmidhuber, “Framewise phoneme classification with bidirectional lstm and other neural network architectures,” *Neural Networks*, vol. 18, no. 5, pp. 602–610, 2005.
[36] F. Eyben, S. Böck, B. Schuller, and A. Graves, “Universal onset detection with bidirectional long short-term memory neural networks.” in *ISMIR*, 2010, pp. 589–594.
[37] M. Schuster and K. K. Paliwal, “Bidirectional recurrent neural networks,” *Signal Processing, IEEE Transactions on*, vol. 45, no. 11, pp. 2673–2681, 1997.
[38] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
[39] P. J. Werbos, “Back-propagation through time: what it does and how to do it,” *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1550–1560, 1990.
[40] A. Robinson and F. Fallside, *The utility driven dynamic error propagation network*. University of Cambridge Department of Engineering,1987.
[41] J. Schmidhuber, “Long short-term memory: Tutorial on lstm recurrent networks,” http://people.idsia.ch/~juergen/lstm/ index.htm, 2003, online resource.
[42] E. Foxley, “Nottingham dataset,” http://ifdo.ca/~sey-mour/nottingham/ nottingham.html, 2011, accessed: 04-19-2015.
[43] D. Temperley, *The Cognition of Basic Musical Structures*. Cambridge: MIT Press, 2001.
[44] M. S. Cuthbert and C. Ariza, “Music21: A toolkit for computer-aided musicology and symbolic music data,” 2010.
[45] M. Greentree. (1996) http://www.jsbchorales.net/index.shtml. Accessed: 04-19-2015.

A Cognitively-Informed Utility for Search and Discovery in Musical Corpora

James Symons[1]
[1]Department of Music Studies, Northwestern University, Evanston, IL, USA

In just the past decade, a host of new technologies have become available for studying musical corpora (such as the Music21 Toolkit and the MATLAB Midi toolbox), that allow us to ask questions and discover features in music that would otherwise be too difficult or time consuming to carry out using traditional methods.

This talk offers a new utility toward this effort, in particular a standalone piece of software, for both discovering phrase-level patterns, including voice-leading schemas in music, and a robust search utility specifically designed to identify schematic-type information in musical corpora. This current project is distinct from other work, in that the discovery and search engine presented here were designed to be as cognitively informed as practically possible, incorporating ideas from schema theory (Gjerdingen 2007), corpus linguistics (Stefanowitsch and Gries, 2004), and computer vision (Marr, 1982), which brings with it interesting consequences. For one thing, the application can be used to reveal patterning in music at various structural scales and can identify similarity in music despite the presence of even a great deal of variety at the musical surface. In this way, the algorithm performs search and finds correspondences that may more closely align with the intuitions of human listeners.

The goal of this talk is to demonstrate the numerous applications of this utility as a tool for music studies, and to discuss relevant music theoretic and cognitive issues that can be investigated using the application. In particular, I demonstrate how these ideas apply to theories of contour, such as those proposed by Morris (1993) and Quinn (1997), and address the problem of equivalence using methods adopted from applied mathematics as an alternative to set theory.

Principal Components Analysis of Musicality in Pitch Sequences

Richard Randall and Adam S. Greenberg

Abstract**—**Musicality can be thought of as a property of sound that emerges when specific organizational parameters are present. We hypothesize that this property is not binary (where an auditory object *is* or *is not* a musical object), but rather exists on a continuum whereby some auditory objects may be considered more or less musical than other auditory objects. We suggest that identification of an auditory object as being more musical than another begins with a modularized analysis of features that coheres into a holistic interpretation. To explore this, we designed two experiments. In the first, 30 subjects evaluated 50 ten-tone sequences according to how musical they thought they were. A special stimulus set was designed that controlled for timbre, pitch content, pitch range, rhythm, note and sequence length, and loudness. Mean z-scored stimulus ratings showed significantly distinct groupings of musical versus non-musical sequences. In the second, a Principal Component Analysis (PCA) of the ratings yielded three components that explain a statistically significant proportion of variance in the ratings. The stimuli were analyzed in terms of parameters such as key correlation, range, and contour. These values were correlated with the eigenvalues of the significant PCA components in order to determine the dominant strategies listeners use to make decisions about musicality.

Keywords**—**Musicality, principal component analysis, auditory perception.

I. Introduction

As a universal trait, humans are born with auditory predispositions that develop over time into musical knowledge and procedures. Adults with no formal musical training have developed the ability to make sophisticated judgments about music through years of exposure to a highly stable and organized auditory environment. Music cognition has been described as an interaction of bottom-up and top-down processes [1]. Auditory scene analysis constructs coherent objects from complex scenes, called auditory objects: the fundamental perceptual unit in audition [4][2][13]. When produced by a single source, auditory objects comprise multiple acoustic events that cohere to form a single auditory stream [27][28]. Stream formation is one of the key features of auditory scene analysis and is the mechanism by which we are able to attend to a specific speaker in a loud crowd (i.e. "the cocktail-party effect" [5]) or focus on a single instrument in a musical ensemble. Once a stream is formed, cognitive functions such as attention and memory are engaged to guide complex behaviors, allowing auditory objects to be processed by domain-specific processes of language or music [14][42]. This paper explores the idea that musicality is a property of auditory objects that emerges when certain organizational features are presented in specific ways. The perception of these musical objects requires multiple psychological processes to evaluate complex acoustic information. Moreover, this work address assumptions that musicality is a property that acoustic information either has or does not have.

Musical memory allows us to recall details about musical objects (e.g., recognition of a familiar melody), and contains schema that represent generalized information about our musical environment. In [1], Bharucha points out that humans have hierarchical schematic representations of musical features, such as metric and tonal organization. The temporal unfolding of a musical object allows its features to be matched with these schematic representations in an automatic way. Repeated exposure to a melody reinforces general mental representations: the stronger the representation, the easier melody recognition becomes [19]. Studies examining deficits in neurologically-impaired individuals observe that music perception is modular for two reasons. First, music is functionally distinct from language [35] and second, it relies on specialized modules that represent distinct music-cognitive processes [34]. Peretz and Coltheart [34] identified distinct neural networks associated with mid-level music processes such as meter, contour, interval, and tonal encoding. Only after an auditory object is processed by these modules can it engage what is referred to as the musical lexicon [34], which contains all of the information about music that one has been exposed to over one's lifetime. The recognition of familiar tunes, they suggest, is dependent on a selection procedure that takes place in the musical lexicon after an auditory object has been determined to be potentially musical by virtue of mid-level analysis of the aforementioned modular features. This idea was also explored in [40], which investigated the musical features that contributed to the recognition of familiar songs, supporting the idea that a song is a temporally unfolding perceptual object. Schulkind, et al. [40] found that familiar song recognition is a holistic process supported by multiple, interacting musical features. Their study was concerned with understanding what musical features correlate with the temporal position of object recognition. Our study is not concerned with familiar-object perception, but rather investigates subjects' perception of musicality in novel (never heard before) sequences.

We hypothesize that the perception of musicality is not binary (where an auditory object *is* or *is not* a musical object),

This research is supported by a Rothberg Research Award in Human Brain Imaging and by NIH grant T32-MH19983.

R. Randall is with the School of Music and the Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, PA 15213 USA (412-268-4785; email: randall@cmu.edu).

A. S. Greenberg is with the Department of Psychology, University of Wisconsin-Milwaukee, Milwaukee, WI 53201 (email: agreenb@uwm.edu).

but rather exists on a continuum whereby some auditory objects may be considered more or less musical than other auditory objects. We suggest that identification of an auditory object as being more musical than another begins with a modularized analysis of features. The results of this analysis produce a percept that may be strongly or weakly associated with memories in the musical lexicon. Musicality, therefore, can be thought of as a property of sound that emerges when certain organizational features are presented in specific ways. By first probing an auditory object's degree of musicality and, second, investigating the features and processes that give rise to the degree of musicality, we can determine how listeners perceive an auditory object as musical.

Music theorists, empirical musicologists, and music psychologists have approached feature identification and analysis in a number of ways. One way is to examine structural features of existing music with the premise that, if a particular feature is significantly present in a specific way, then there is a mental representation or process that has developed in response to this feature. Another approach is to posit a general idea about "what music is" from the experimenter's experience. In both cases, features lead to testable hypotheses and the importance of the feature is established when observable behavior can be modulated by the manipulation of that feature. In both approaches, however, an auditory-object's status as "music" is taken *a priori*. While the assumption that Mozart's "Hunt Quartet" or an ethnic folk song is *axiomatically* music is appropriate in some contexts, it can inhibit us from understanding what a subject-oriented mental representation of music might look like. In the current study we seek to uncover an operational definition of music by first observing subjects' behavior and then examining the stimuli in an attempt understand how stimuli modulate behaviors. To test this, we asked subjects to listen to randomly generated pure-tone sequences and then rate each sequence on its musicality. Our first hypothesis is that subjects will be able to effectively group these sequences according to their musicality. Our second hypothesis is that a measure can be devised to quantitatively describe the psychological processes involved in judging the musicality of simple auditory stimuli. We created a profile for each sequence, comprised of a set of metrics that describe structural features commonly discussed in the music theory and music psychology literature. A principal component analysis (PCA) was performed on the ratings and the resulting components were correlated with the profiles to understand which features best account for the variance found in the musicality ratings. This approach allows for an exploration of the low-level auditory features that give rise to the perception of musicality in auditory objects. The following describes our two experiments and discussion is reserved for the end.

II. Experiment 1

A. Methods

1) Subjects

30 participants (17 female, 13 male) were recruited using the Carnegie Mellon University Center for Behavioral and Decision Research subject pool. Ages ranged from 19 to 58 years old with mean of 30.26 (*SD* 13). All subjects self-identified as having normal hearing and normal or corrected-to-normal vision. All were considered non-musicians with a mean 2.45 years of formal music training. All were native speakers of American English. No subjects reported having absolute pitch (AP) or knowledge of any family members with AP.

2) Stimuli

We used 50 sequences, each with 10 pure tones of 500ms duration, with an inter-stimulus-interval of 500ms. Each of the 10 tones was chosen randomly from the diatonic collection corresponding to the G-major scale (G4-F#5 or 392Hz-740Hz). Sequences were randomly transposed to all 12 pitch-class levels and then transposed back to the original G4-F#5 span creating an equal distribution of pitches across the chromatic scale.

3) Task

Subjects were asked to provide a rating of musicality for each sequence. They were instructed to rate each sequence on a Likert scale of 1 to 5. If they thought the sequence was very musical, they were asked to press the '5' key on the keyboard; if they thought the sequence was not musical at all, they were asked to press the '1' key on the keyboard. Subjects were asked to use the entire range of ratings and were given 1.5 seconds to respond. If they responded before the sequence had completed or failed to respond in 1.5 seconds, the response was not recorded and the experiment moved on to the next trial.

4) Design

Subjects completed a practice block of six trials after which they were asked if they understood the task. Subjects were given the option to complete additional practice trials or to move on to the main part of the experiment. The experiment was comprised of 15 blocks, each with 20 trials. 50 sequences were played eight times apiece for a total of 400 trials. Trials were presented in 20 blocks, with a forced 20-second break between each block. Stimuli were presented via headphones (Sennheiser HD210) at a fixed, comfortable listening volume (~82dBA-SPL). The paradigm ran on an Apple Mac Mini and responses were recorded on a standard computer keyboard. The paradigm was coded using MATLAB with Psychtoolbox extensions [3][33].

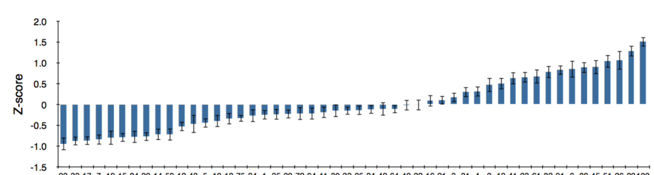


Fig. 1 Group (N=30) z-scored stimulus ratings of 50 sequences. Lower scores indicate less-musical sequences and higher scores indicate more-musical sequences.

5) Results

The results, shown in Fig. 1, show clear separability between the most musical and the least musical sequences, as well as a graded representation across the scale. Subjects used the entire scale in rating the sequences. For analysis, ratings were divided into two equal sized groups (lower 25 and upper

25). An ANOVA (alpha = 0.05) shows a significant difference between the two groups, $F(21,3)=17.45$, $p=0.019$. Fig. 2 shows the three most musical and the three least musical sequences for reference. This result demonstrates that the perception of musicality is inter-subjectively stable and is a quality that varies across stimuli.

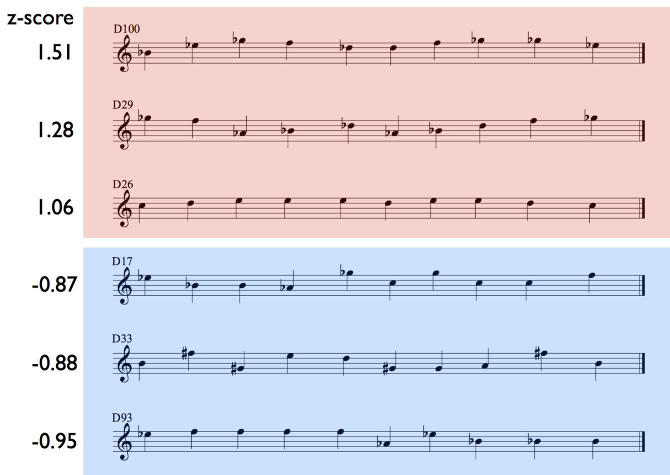



Fig. 2 The three-most (red) and three-least (blue) musical sequences with z-scores.

III. Experiment 2

A. Methods

We identify five feature categories that figure predominantly in the music-theoretic and music-psychological literature: interval, contour, tonality, motive, and entropy. This list is not complete, nor comprehensive, but it is representative. Each feature is represented as a metric so we can understand the consequences of its variation. Metrics were derived from each sequence using *Music21* [7], *Midi Toolbox* [10], standard statistical analysis software (MATLAB and SPSS), and custom scripts. Like *Humdrum Toolkit* [15] before it, *Music21* and *Midi Toolbox* come with a set of analytic tools or routines grounded in recent psychological and music-theoretic literature that effectively explore certain musical features. Other features we explored required more creative treatments for them to be useful in our analysis. Below are descriptions of the features we used, their psychological and music theoretical context, and some general predictions about how they might relate to the results of experiment 1.

1) Intervals

A pitch interval is the distance in semitones between two tones and a melodic interval is the distance between two adjacent pitches in a sequence. We created an intervallic profile for each sequence by calculating the mean interval size (the average of all consecutive melodic intervals), standard deviation (SD) of the mean, and range [19]. In addition, we calculated the melodic interval variability (MIV), a version of the coefficient of intervallic variation of a sequence used in music/language research [32].

2) Contour

Contour refers to the shape of musical materials (e.g., pitches, rhythms, timbres, tempi,) [29]. In a simple melodic context where variation occurs only in the pitch domain, contour refers to shape of the sequence as the pitch ascends and/or descends in frequency space. Analysis of the Essen Folk Song collection reveals that the predominant contour for 10-note sequences is a clear arch shape that gradually rises from a starting point, peaks midway, and returns to end at the original or near-original starting point [16]. Using [16]'s 10-note arch as an archetype, we defined a function that returns a value that represents how well each sequence fits this archetype (e.g. correlation). The contour correlation value ranges from 1 to 0, where 1 represents equivalence between a sequence and the archetype. The contour of a sequence is usually easier to remember than exact interval information [9][8]. Given the strength of Huron's findings, we predict that sequences that more closely fit this archetype will be perceived as being more musical than sequences that do not fit the archetype.

3) Tonality

Tonality is the property whereby tones in a scale are hierarchically organized around a central pitch. Key finding algorithms correlate a musical excerpt with a key according to a probabilistic distribution based on the Krumhansl tonal profile [22]. Supporting [34]'s findings, recent behavioral research asserts that identification of a tonal center is an elemental process at the core of how all listeners experience music [11]. We used the key finding algorithms implemented in *Music21* to find key centers of each sequence [7]. While the algorithms are in many ways similar, each is methodologically different and the methods described by [7] are not comprehensive with respect to all available approaches. Since the sequences we used are only 10 notes in length and the key-finding algorithms are based on statistics whose power increases with the number of pitches, we selected only the highest correlation for each sequence. The actual key correlation of a sequence is irrelevant for this study; our *maximum key correlation* metric represents the likelihood that a certain key, or tonal center, can be inferred. Since [34] proposes that the inference of key is necessary for an auditory object to be musical, we hypothesize that sequences with higher key correlations will be considered more musical than those with lower correlations.

4) Motive

The use of motives (small, related, and easily remembered musical ideas) is an important component of the overall musical experience, as motives allow listeners to conflate smaller musical ideas into a single larger concept. Following [4], when sounds cohere they form an auditory stream that is separable from other sounds or other streams. However, [49] maintains that coherence allows multiple streams to be compared with one another, thus promoting a higher-level of perceptual organization – a property [49] shows is commonly found in music of the Western classical canon. In this sense, motivic coherence is similar to compression, whereby complex auditory scenes comprising multiple auditory objects are made easier to interpret and remember by virtue of a similarity between objects. While there is a literature addressing the segmentation of music into smaller units, such approaches rely on features (e.g., such as varied rhythm) that our melodic stimuli do not possess. Because our sequences are relatively feature-poor, we used prediction by partial matching (PPM) to measure the compressibility of each sequence. PPM is an adaptive statistical compression technique that predicts the *n*th symbol in a string by using a context that optimally varies in length according to its predictive success [26]. We used a PPM encoder on consecutive intervals in each sequence and limited the context size to a maximum of 5. Since motive-rich music is more highly valued than motive-poor music by some music scholars [31][6][12][49], we expect that greater compressibility will correlate with greater musicality.

5) Entropy

Probabilistic entropy has a rich history in music scholarship largely centered on style analysis [48][20][41]. It has also been used to model melodic complexity, musical expectation, and aesthetic experience [25][10][23]. Entropy is closely related to ideas of compression found in our PPM analysis and the probabilistic distribution found in our key finding analysis. Relative entropy (Hr) is a convenient and standard metric for characterizing how varied the distribution of symbols (pitches) is given a fixed alphabet (the diatonic scale). It is difficult to make predictions about how varied levels of uncertainty might affect musicality. We might predict that a 10-note sequence composed using a single pitch might be equally as musical as a 10-note sequence employing all seven diatonic pitches. Nevertheless, because of its continued appearance in music-related studies, we included relative entropy in the sequence profile.

Fig. 3 shows the aforementioned metrics associated with the three most- and least-musical sequences.

D93	D33	D17		D26	D29	D100
-0.95	-0.88	-0.87	Z-Score	1.06	1.28	1.51
9	10	10	Range	4	10	8
3.34	5.56	4.44	Mean	1.33	2.36	1.76
2.56	3.43	3.06	SD	0.94	3.33	2.33
1.31	0.62	0.62	MIV	0.71	0.71	0.76
0.71	0.82	0.56	Key	0.64	0.82	0.71
0.67	0.82	0.71	Contour	0.57	0.71	0.49
0.65	0.86	0.78	Hr	0.48	0.81	0.97
43.81	64.48	53.97	PPM	36.98	54.53	51.68

Least Musical ← → Most Musical

Fig. 3 Feature metrics for the three-most and three-least musical sequences shown.

IV. Results

The first three intervallic features (Range, Mean, and SD) are strongly negatively correlated (p=<0.001) with z-mean scores. Subjects found that sequences with smaller range, smaller mean-interval size, and smaller standard deviation of the mean to be more musical (Table 1). This confirms our prediction and supports the Huron's detailed work showing that cross-culturally, sequences privilege small intervals [19]. Interestingly, features such as MIV, Contour, Key, Hr, and PPM do not significantly correlate with the ratings.

We hypothesize that subjects' ratings will have uncorrelated principal components that will each correlate with the different musical features described above. That is, components are separable in terms of these features suggesting that these features are used, to varying degrees, in combination when subjects make musicality judgments of pitch sequences. PCA of the ratings was performed and permutation testing showed that only the first three components were significant, explaining a combined 38% of the variance; however, this is a conservative measure. Because information is often embedded in additional components, we include the top five. The feature profile of each tune was correlated with the eigenvalues of the significant PCA components in order to identify dominant strategies listeners use to make decisions about musicality. The PCA analysis is shown in Fig. 4.

Table 1
Significant correlations between features and ratings.

		Zscore
Zscore	Pearson Correlation	1
	Sig. (2-tailed)	
	N	50
Mean	Pearson Correlation	-.619**
	Sig. (2-tailed)	.000
	N	50
Range	Pearson Correlation	-.491**
	Sig. (2-tailed)	.000
	N	50
SD	Pearson Correlation	-.697**
	Sig. (2-tailed)	.000
	N	50

Table 2 shows correlations between eigenvalues of the first five components and the eight metrics. Significant correlations are highlighted in red. Of the eight, only intervallic features of Range, Mean, and SD correlate significantly with the first component. Mean also correlates strongly with the second component as does Key. Hr and PPM appear less strongly followed by MIV. Contour correlates with the third component followed closely by a reappearance of PPM.

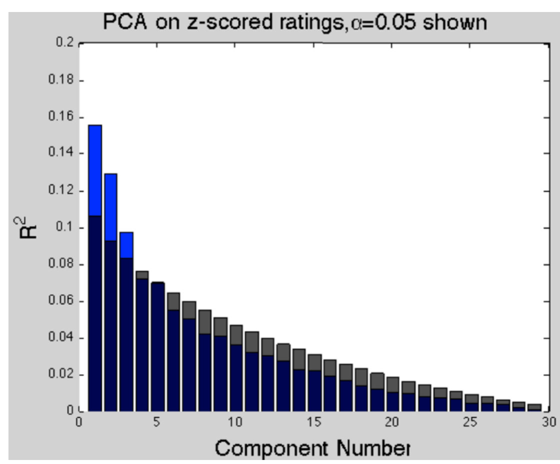


Fig. 4 A principal components analysis (PCA) on subjects' ratings returns three significant uncorrelated components explaining 38% of the variance. Gray bars indicate critical R^2 values from permutation testing; blue bars indicate R^2 values from PCA.

V. Discussion

The goal of this study was to understand the conditions under which randomly generated sequences are organized into (or perceived as) auditory objects. Despite the extensive work that has been done on auditory stream formation, few studies have examined how this process plays out in novel musical contexts. Notable treatments address how bistable auditory streams are formed in polyphonic music and how acoustic features, such as timbre, contribute to object formation [47][24][45]. Unlike previous studies that present multi-stable auditory scenes, this study presents a single stream in a non-competitive auditory environment. The streams were controlled for range, timbre, tempo, and event duration, so as to minimize the possibility that any one stream could be interpreted as multi-stable. However, the perception of multiple streams and how it contributes to musical object formation remains an open question.

TABLE II
SIGNIFICANT CORRELATIONS (IN RED) BETWEEN FEATURES AND COEFFICIENTS

		Comp1	Comp2	Comp3	Comp4	Comp5
Mean	Pearson Correlation	.647**	.445**	-.028	-.076	.001
	Sig. (2-tailed)	.000	.001	.844	.598	.995
	N	50	50	50	50	50
Range	Pearson Correlation	.576**	.170	.233	.106	-.048
	Sig. (2-tailed)	.000	.237	.104	.463	.741
	N	50	50	50	50	50
SD	Pearson Correlation	.652**	.170	-.009	.230	.205
	Sig. (2-tailed)	.000	.239	.952	.108	.152
	N	50	50	50	50	50
Key	Pearson Correlation	.033	.445**	.039	-.168	.026
	Sig. (2-tailed)	.822	.001	.787	.244	.860
	N	50	50	50	50	50
Hr	Pearson Correlation	.246	.412**	.262	.012	-.190
	Sig. (2-tailed)	.085	.003	.066	.936	.186
	N	50	50	50	50	50
PPM	Pearson Correlation	.093	.358*	.451**	.065	-.267
	Sig. (2-tailed)	.521	.011	.001	.652	.061
	N	50	50	50	50	50
MIV	Pearson Correlation	-.042	-.327*	-.005	.286*	.293*
	Sig. (2-tailed)	.771	.021	.973	.044	.039
	N	50	50	50	50	50
Contour	Pearson Correlation	.060	-.244	.494**	.033	-.134
	Sig. (2-tailed)	.677	.087	.000	.820	.354
	N	50	50	50	50	50

Additional analysis was performed in an attempt to understand how select musical features interact with each other to support the perception of musicality. Fig. 5 shows the linear combination of metrics that significantly correlate with component 1 plotted against the eigenvalues. Musicality ratings are color coded with red as most musical and blue as least musical. A clear pattern is observed whereby higher musicality ratings cluster towards smaller values and lower musicality ratings cluster upward toward larger values. This is shown more clearly in Fig. 6, which limits the linear combination to significantly correlated metrics of the ten most and least musical.

Comparing Figs. 5 and 6 we see that there is an organizational difference between high and low musical sequences. For the significant metrics, their strong bilateral separation between high and low musical sequences along the x axis shows that subjects are using these features (Mean, Range, and SD) consistently to make musicality judgments. Furthermore, separable groups along the y axis show that as composite values of the metrics increase, so too does variance.

In contrast, Fig. 7 plots the linear combination of metrics that *do not* significantly correlate with component 1 against its eigenvalues. The pattern is more diffuse with a slight bilateral distribution of high versus low musicality across the x axis, but no grouping along the y axis. We interpret this as evidence that the non-significant metrics (Key, Hr, PPM, MIV, and Contour) are not contributing meaningfully to the variance at this level.

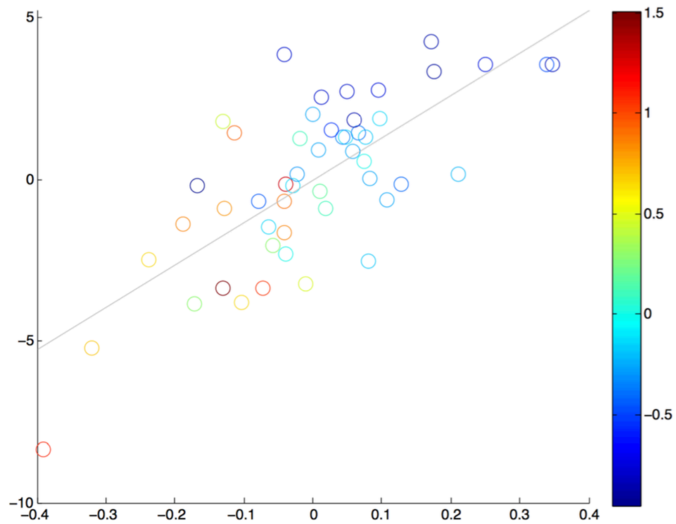


Fig. 5 Linear combination of metrics significantly correlated with component 1 plotted against eigenvalues. Z-score for each sequence is color-coded (red = high, blue=low).

Our analysis shows that common features used to describe musical works can explain how subjects determine the degree to which an auditory object is musical in a limited way. It raises interesting questions about the ontological role of these (and other) analytic descriptions. For instance, we might claim that while key membership is important for determining whether or not a sequence of tones is a *musical* sequence of tones, that quality might only be important if specific intervallic features are present as well. We might also assert that while contours in music exhibit regular patterns (e.g., the "melodic arch"), such patterns may not play a particularly strong role in musical auditory object formation, or the role they play might be dependant on the presence other features.

PCA has been used by [39] and [37] in music perception studies to simplify high-dimensional models. Here, we use PCA in an exploratory way to identify components that explain the variance found in the data. All features described in section III-A significantly correlate with one or more of the components. However, components, even if significant, may not be statistically independent from one another; therefore, we cannot make any claims about the independence of components: nor can we make claims about ordering. In other words, the fact that component 1 is most strongly correlated with interval analysis does not mean that it does not impact component 2.

Music perception in its most basic sense is dependent on listeners' ability to recognize the presence or absence of structural features and to build a representation of auditory objects. This study explores how such features relate to the perceived musicality of a pitch sequence. It combines perspectives from music theory, computational musicology, and music perception to identify a set of features that, while not exhaustive, all represent musical experiences. Experiment 2 is constrained by the set of features we have chosen. Whether or not a feature has explanatory power, of course, depends on the encoding of that feature. It is possible that we are limited in thinking about musical features in specific ways, finding just enough evidence to continue considering them an important part of our mental representations of music.

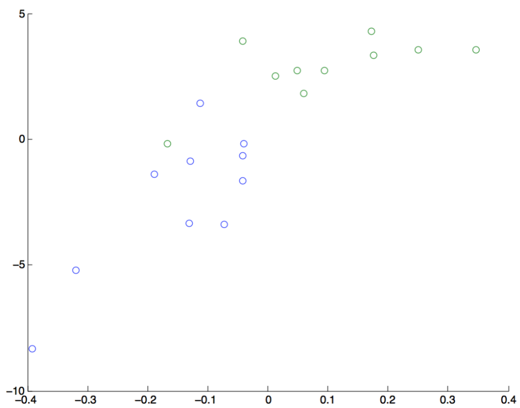


Fig. 6 Linear combination of metrics significantly correlated with component 1 plotted against eigenvalues for the 10 most musical (blue) and 10 least musical (green) sequences.

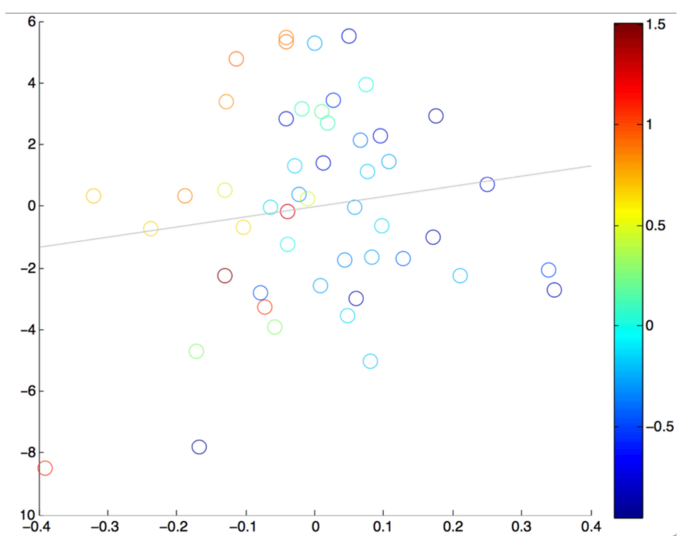


Fig. 7 Linear combination of metrics not significantly correlated with component 1 plotted against eigenvalues. Z-score for each sequence is color coded (red = high, blue=low).

Our approach to explicate these mental representations provides a new way to examine music. This study sought to understand the boundary conditions for how auditory objects can become musical objects. We show that musicality is a variable quality that auditory objects possess in greater or lesser degrees. We also show that there is considerable intersubjective agreement about what constitutes a highly musical object. We see that musicality is a property of auditory objects that emerges under specific conditions, and exists on a continuum. Placement on this continuum is the result of multiple interacting features. Our results are informative for future music studies and our empirically derived collection of 10-tone sequences is fertile territory for exploration. Future work includes identifying the neural correlates of the varied perception of auditory objects.

References

[1] J. J. Bharucha, "Music cognition and perceptual facilitation: A connectionist framework," *Music Perception*, pp. 1–30, 1987.

[2] J. K. Bizley and Y. E. Cohen, "The what, where and how of auditory-object perception," *Nature Reviews Neuroscience*, vol. 14, no. 10, pp. 693–707, 2013.

[3] D. H. Brainard, "The psychophysics toolbox," *Spatial vision*, vol. 10, pp. 433–436, 1997.

[4] A. S. Bregman, *Auditory scene analysis: The perceptual organization of sound*. MIT press, 1994.

[5] E. C. Cherry, "Some experiments on the recognition of speech, with one and with two ears," *The Journal of the acoustical society of America*, vol. 25, no. 5, pp. 975–979, 1953.

[6] E. T. Cone, "On derivation: Syntax and rhetoric," *Music Analysis*, vol. 6, no. 3, pp. 237–255, 1987.

[7] M. S. Cuthbert and C. Ariza, "Music21: A toolkit for computer-aided musicology and symbolic music data," 2010.

[8] W. J. Dowling and D. S. Fujitani, "Contour, interval, and pitch recognition in memory for melodies," *The Journal of the Acoustical Society of America*, vol. 49, no. 2B, pp. 524–531, 1971.

[9] W. J. Dowling, "Scale and contour: Two components of a theory of memory for melodies.," *Psychological review*, vol. 85, no. 4, p. 341, 1978.

[10] T. Eerola and P. Toiviainen, "MIDI toolbox: MATLAB tools for music research," 2004.

[11] M. M. Farbood, G. Marcus, and D. Poeppel, "Temporal dynamics and the identification of musical key.," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 39, no. 4, p. 911, 2013.

[12] R. O. Gjerdingen, "Revisiting Meyer's' grammatical simplicity and relational richness,'" 1992.

[13] T. D. Griffiths and J. D. Warren, "What is an auditory object?," *Nature Reviews Neuroscience*, vol. 5, no. 11, pp. 887–892, 2004.

[14] L. L. Holt and A. J. Lotto, "Speech Perception Within an Auditory Cognitive Science Framework," *Current Directions in Psychological Science*, vol. 17, no. 1, pp. 42–46, Feb. 2008.

[15] D. Huron, "The Humdrum Toolkit: Software for Music Research," *Center for Computer Assisted Research in the Humanities, Ohio State University, copyright*, vol. 1999, 1993.

[16] D. Huron, "The melodic arch in Western folksongs," *Computing in Musicology*, vol. 10, pp. 3–23, 1996.

[17] D. Huron, "Music research using Humdrum: A user's guide," *Center for Computer Assisted Research in the Humanities, Stanford, California*, 1999.

[18] D. Huron, "Music information processing using the Humdrum Toolkit: Concepts, examples, and lessons," *Computer Music Journal*, vol. 26, no. 2, pp. 11–26, 2002.

[19] D. B. Huron, *Sweet anticipation: Music and the psychology of expectation*. MIT press, 2006.

[20] L. Knopoff and W. Hutchinson, "Entropy as a measure of style: The influence of sample length," *Journal of Music Theory*, vol. 27, no. 1, pp. 75–97, 1983.

[21] C. L. Krumhansl and M. A. Schmuckler, "Key-finding in music: An algorithm based on pattern matching to tonal hierarchies," in *Mathematical Psychology Meeting*, 1986.

[22] C. L. Krumhansl, "Perceptual structures for tonal music," *Music Perception*, pp. 28–62, 1983.

[23] E. H. Margulis and A. P. Beatty, "Musical style, psychoaesthetics, and prospects for entropy as an analytic tool," *Computer Music Journal*, vol. 32, no. 4, pp. 64–78, 2008.

[24] S. McAdams, "Perspectives on the contribution of timbre to musical structure," *Computer Music Journal*, vol. 23, no. 3, pp. 85–102, 1999.

[25] L. B. Meyer, *Emotion and meaning in music*. University of chicago Press, 2008.

[26] A. Moffat, "Implementing the PPM data compression scheme," *Communications, IEEE Transactions on*, vol. 38, no. 11, pp. 1917–1921, 1990.

[27] B. C. Moore and H. Gockel, "Factors influencing sequential stream

segregation," *Acta Acustica United with Acustica*, vol. 88, no. 3, pp. 320–333, 2002.
[28] B. C. Moore and H. E. Gockel, "Properties of auditory stream formation," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 367, no. 1591, pp. 919–931, 2012.
[29] R. Morris, "Composition with Pitch-classes," *Yale University*, 1987.
[30] R. D. Morris, "New directions in the theory and analysis of musical contour," *Music Theory Spectrum*, vol. 15, no. 2, pp. 205–228, 1993.
[31] E. Narmour, *The analysis and cognition of melodic complexity: The implication-realization model*. University of Chicago Press, 1992.
[32] A. D. Patel, *Music, language, and the brain*. Oxford university press, 2010.
[33] D. G. Pelli, "The VideoToolbox software for visual psychophysics: Transforming numbers into movies," *Spatial vision*, vol. 10, no. 4, pp. 437–442, 1997.
[34] I. Peretz and M. Coltheart, "Modularity of music processing," *Nature neuroscience*, vol. 6, no. 7, pp. 688–691, 2003.
[35] I. Peretz and K. L. Hyde, "What is specific to music processing? Insights from congenital amusia," *Trends in cognitive sciences*, vol. 7, no. 8, pp. 362–367, 2003.
[36] I. Peretz, "The nature of music from a biological perspective," *Cognition*, vol. 100, no. 1, pp. 1–32, 2006.
[37] L. Quinto, W. F. Thompson, and A. Taylor, "The contributions of compositional structure and performance expression to the communication of emotion in music," *Psychology of Music*, p. 0305735613482023, 2013.
[38] H. Schaffrath and D. Huron, "The Essen folksong collection in the humdrum kern format," *Menlo Park, CA: Center for Computer Assisted Research in the Humanities*, 1995.
[39] E. G. Schellenberg, "Simplifying the implication-realization model of melodic expectancy," *Music Perception: An Interdisciplinary Journal*, vol. 14, no. 3, pp. 295–318, 1997.
[40] M. D. Schulkind, R. J. Posner, and D. C. Rubin, "Musical Features That Facilitate Melody Identification: How Do You Know It's '"Your"' Song When They Finally Play It?," *Music Perception: An Interdisciplinary Journal*, vol. 21, no. 2, pp. 217–249, 2003.
[41] J. L. Snyder, "Entropy as a measure of musical style: the influence of a priori assumptions," *Music Theory Spectrum*, vol. 12, no. 1, pp. 121–160, 1990.
[42] D. L. Strait, N. Kraus, A. Parbery-Clark, and R. Ashley, "Musical experience shapes top-down auditory mechanisms: evidence from masking and auditory attention performance," *Hearing research*, vol. 261, no. 1, pp. 22–29, 2010.
[43] D. Temperley, "What's key for key? The Krumhansl-Schmuckler key-finding algorithm reconsidered," *Music Perception*, pp. 65–100, 1999.
[44] D. Temperley, *Music and probability*. MIT Press Cambridge^ eMass Mass, 2007.
[45] M. Uhlig, M. T. Fairhurst, and P. E. Keller, "The importance of integration and top-down salience when listening to complex multi-part musical stimuli," *Neuroimage*, vol. 77, pp. 52–61, 2013.
[46] N. L. Wood and N. Cowan, "The cocktail party phenomenon revisited: attention and memory in the classic selective listening procedure of Cherry (1953).," *Journal of Experimental Psychology: General*, vol. 124, no. 3, p. 243, 1995.
[47] J. K. Wright and A. S. Bregman, "Auditory stream segregation and the control of dissonance in polyphonic music," *Contemporary Music Review*, vol. 2, no. 1, pp. 63–92, 1987.
[48] J. E. Youngblood, "Style as information," *Journal of Music Theory*, vol. 2, no. 1, pp. 24–35, 1958.
[49] L. M. Zbikowski, "Musical coherence, motive, and categorization," *Music Perception: An Interdisciplinary Journal*, vol. 17, no. 1, pp. 5–42, 1999.

Comparison of Tonality Models in Measuring Chord Sequence Similarity

Saebyul Park, Jeounghoon Kim, Juhan Nam
GSCT, KAIST, Republic of Korea
saebyul_park@kaist.ac.kr, miru@kaist.ac.kr, juhannam@kaist.ac.kr

ABSTRACT

The objective of this study is to examine chord sequence similarity measures and experimentally assess their relationship with human perception of the similarity. We present five different types of chord sequence similarity measures based on different tonality models, including 1) tonal dissonance of intervals, 2) circle of fifths, 3) harmonic relations, 4) tonal pitch space, and 5) hierarchy of harmonic stability. For the evaluation, we collected 50 chord sequence pairs from US musical copyright infringement cases. Also, we surveyed human evaluation to compare it with the computational chord sequence similarity. The results show that those based on tonal pitch space and hierarchy of harmonic stability are relatively more correlated with the human judgement data and a combination of the two similarity measures further increases the correlation.

I. INTRODUCTION

Measuring music similarity has always received high attention in the fields of music information retrieval (MIR) and music psychology research, since classification and analysis of music basically start by concentrating on how individual songs share similar musical characteristics. Especially, measuring music similarity based on *harmony*, which is one of the core features of western tonal music, has great advantages in plagiarism detection, genre classification, identifying cover song and music recommendation systems (Casey et al., 2008).

Recently, various research on harmonic similarity has been carried out in the form of comparing symbolic chord sequences (B. De Haas, Veltkamp, & Wiering, 2008; W. B. De Haas, Wiering, & Veltkamp, 2013; Freedman, 2015; Hanna, Robine, & Rocher, 2009; Rocher, Robine, Hanna, & Desainte-Catherine, 2010). Even though these studies made great contributions in systematically and quantitatively measuring chord sequence similarity encompassing music theoretical models, we believe that there are additional musical factors that deserve to be included in the analysis. Moreover, previous work rarely has considered the relationship between computational measures of chord sequence similarity and human perception of the similarity.

In this paper, we present five different types of chord sequence similarity measures based on different tonality models and compare them to human judgements of the similarity. Through the experiment, we evaluate the computational similarity measures and show how they are correlated to the human data.

II. METHODS

A. Study 1: Computational Chord Sequence Similarity

For the experiment, we collected 50 chord sequence pairs extracted from songs in US musical copyright infringement cases as they were regarded as similar, but not identical. Specifically, we normalized the key by transposing all songs to C major. We cut the chord sequences into a set of 4 chords from each of song pairs and limited the chords to triads in major and minor mode (i.e., major triad, minor triad, diminished triad, and augmented triad). In order to compute chord similarity measures, we represented the chords as a triplet of pitch classes in a range of 0 to11. For example, C major triad is represented as “0 4 7”, C minor triad is represented as “0 3 7” and A minor triad is represented as “9 0 4”.

The computational similarity measures are based on edit distance in common. The edit distance, also known as the Levenshtein algorithm, is a metric that computes the minimum number of operations needed to transform one sequence into the other. The operations between sequences include deletion, insertion, and substitution of symbols. Let y be an operation cost function, e the empty string, two chord sequences, $a = a_1 \ldots a_n$ and $b = b_1 \ldots b_n$. Then, the edit distance matrix, d_{mn} between a and b is computed as follows:

$$d_{00} = 0,$$
$$d_{i0} = i \quad for\ 1 \le i \le m,$$
$$d_{0j} = j \quad for\ 1 \le j \le n$$
$$d_{ij} = \begin{cases} d_{i-1,j-1} & for\ a_j = b_i \\ min \begin{cases} d_{i-1,j} + y_i(e \to b_j) \\ d_{i,j-1} + y_d(a_i \to e) \\ d_{i-1,j-1} + y_s(a_j \to b_i) \end{cases} & for\ a_j \ne b_i \end{cases}$$
$$for\ 1 \le i \le m, 1 \le j \le n$$

While fixing both insertion and deletion costs to a constant value of 1, we vary the substitution cost based on distance between the two chords derived from different tonality models. In the following subsections, we describe how to quantify the substitution cost in each of the models.

1) *Tonal dissonance of intervals.* With this model, we use the root of the triads only and thus measure the chord distance (i.e., the substitution cost) from the two root notes. Table 1 shows the dissonance rating of interval (Nordmark & Fahlén, 1988). We compute the substitution cost by taking the number of semitones from 1 to 11 and normalizing the dissonance rating values to the range between 1 and 2.

2) *Circle of fifths* The second substitution cost is based on the circle of fifths, which is a graphical representation of the 12 notes scale placed onto a circle where neighbouring notes are separated by a fifth interval (Figure 1). With this model, we also use the root of the triads only. The chord distance is defined as the smaller number of steps in the circle of fifths either clockwise or counter-clockwise. We compute the substitution cost by normalizing the number of non-zero steps to the range between 1 and 2.

Table 1. Dissonance rating between two intervals (Cook, 1999; Nordmark & Fahlén, 1988)

Interval Name	Number of Semitones	Dissonance Rating(1-7)
Octave	12	1.7
Fifth	7	1.7
Fourth	5	2.0
Major third	4	2.0
Major sixth	9	2.4
Minor third	3	2.6
Minor sixth	8	3.0
Minor seventh	10	3.3
Major second	2	3.9
Tritone	6	4.0
Major seventh	11	5.3
Minor second	1	5.7
Minor ninth	13	5.8

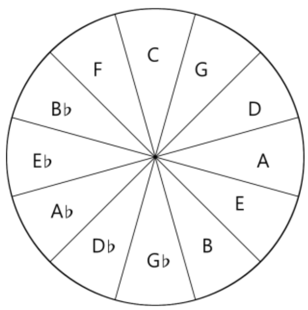


Figure 1. The Circle of fifths

3) *Harmonic relations* The third substitution cost function is defined based on the theoretical harmonic relations (Lerdahl, 2001). The distance between chord pairs, which all belong to the harmonic function in C major as we transposed all chords to this key, are selected from the theoretical harmonic relations (See Table 2). The values between 5 and 8 are scaled to the range between 1 and 2 as a substitution cost.

Table 2. Theoretical Harmonic Relations from Lerdahl (2001)

	Ⅰ	Ⅱ	Ⅲ	Ⅳ	Ⅴ	Ⅵ	Ⅶ
Ⅰ	0						
Ⅱ	8	0					
Ⅲ	7	8	0				
Ⅳ	5	7	8	0			
Ⅴ	5	5	7	8	0		
Ⅵ	7	5	5	7	8	0	
Ⅶ	8	7	5	5	7	8	0

4) *Tonal pitch space* For the fourth substitution cost function, the tonal pitch space (TPS) model (Lerdahl, 2001) was used to calculate a distance between two chords. The basis of the model is the basic space (see Figure 2) which comprises five hierarchical levels (a-e) consisting of pitch class subsets ordered from root level to chromatic level. The chordal distance is calculated by the number of non-common pitch classes within the basic spaces of two chords, divided by two. We calculated the distance from level a to c (level a-c), as we limited our chord selection to triads only, resulting the maximum distance of 6. Again, the non-zero chord distance is scaled to the range between 1 and 2 as a substitution cost.

(a) root level:	0											
(b) fifth level:	0							7				
(c) triadic level:	0				4			7				
(d) diatonic level:	0		2		4	5		7		9		11
(e) chromatic level	0	1	2	3	4	5	6	7	8	9	10	11

Figure 2. Diatonic basic space for C major triad (C=0, C#=1…B=11)

5) *Hierarchy of harmonic stability* For the fifth substitution cost function, we use the distance principle derived from hierarchy of harmonic stability are applied (Bharucha & Krumhansl, 1983). In this model, the psychological distances between chords reflect both key membership and stability within the key: 1) Chords from the same key are perceived as more closely related than those from different keys, 2) Chords in a harmonic core (Tonic, subdominant and dominant) are perceived as more closely related to each other compared to other chords from the key, but not in core. Based on this model, the substitution cost is computed when two different chords C_1 and C_2 are given as followed:

$$
\begin{array}{lll}
cost = 0 & \text{when} & C_1 = C_2 \\
cost = 1 & \text{when} & C_1, C_2 \in K \text{ and } C_1, C_2 \in S \\
cost = 2 & \text{when} & C_1, C_2 \in K \text{ and } C_1 \text{ or } C_2 \notin S \\
cost = 3 & \text{when} & C_1 \; or \; C_2 \notin K,
\end{array}
$$

where K contains the seven triads built upon the seven degrees of the diatonic scale and S is the set containing the three harmonic chords (I, IV, and V) of the harmonic core. The non-zero substitution cost is scaled to the range between 1 and 2 to be used in the edit distance.

B. Study 2: Human Survey Experiment

The objective of the survey experiment is to gather human similarity judgment data to compare the results with the computed similarity values. A total of 28 subjects participated in the experiment. Participants were 24–58 years old (mean μ = 27.7) and were recruited without regard to their musical training.

Chord sequence pairs generated using US copyright infringement cases from study 1 were also used in audio form in study 2. Chords were generated in a C major key, within a range of two octaves centred at middle C. Chord sequences were composed of 4 chords, with a beat of a quarter notes each (140 bpm), resulting the total length of the audio of 6 seconds.

Human experiments were conducted as an on-line survey. Before starting to answer the questionnaire, participants were informed about the task and presented with examples of very similar and very dissimilar chord sequences. During the survey, participants listened to the audio clips for each chord sequence pair, and were asked to rate the similarity of these pairs on a scale from 1 to 4 (1=very similar, 2=similar, 3=dissimilar, and 4=very dissimilar). It took approximately 30 minutes to finish.

III. RESULTS

The comparison results are summarized in Figure 3. Pearson's correlation coefficient was used to compare the results between the computational measures and human survey data.

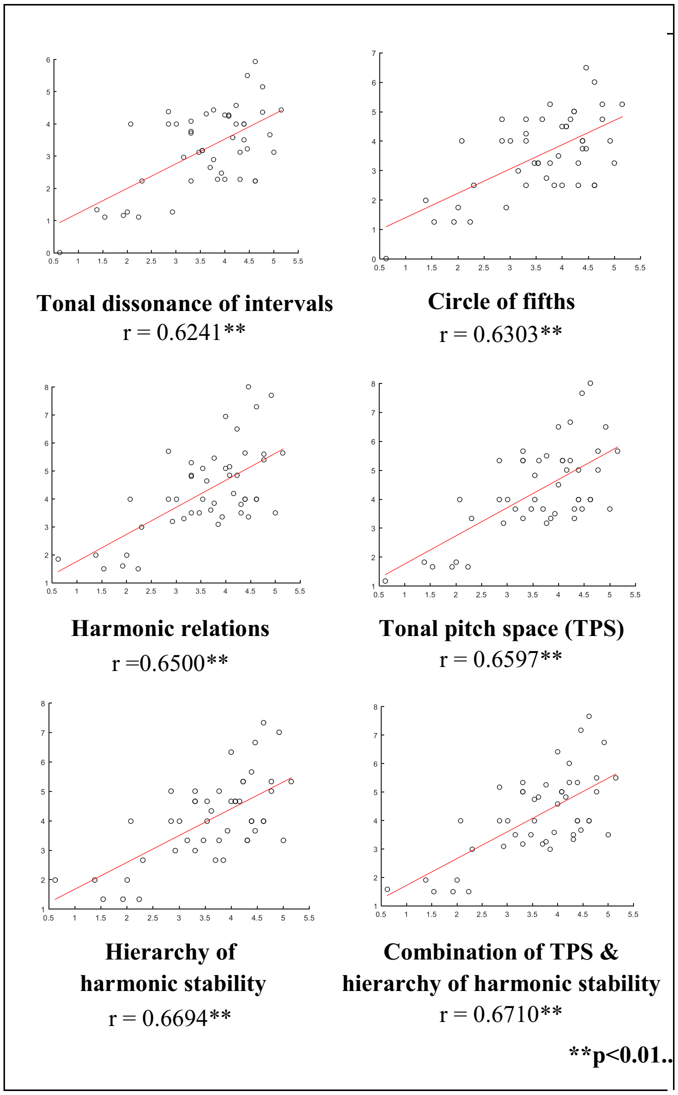


Figure 3. Pearson's correlation coefficient between computational measures and human judgements in chord sequence similarity

As shown in Figure 3, all of computed similarity measurements had significant correlation with human judgment of similarity ($p<0.01$). The highest correlation was achieved by the similarity measure based on Bharucha–Krumhansl's hierarchy of harmonic stability model ($r=0.669$, $p<0.01$), followed by the TPS model ($r=0.659$, $p<0.01$). This suggests that key membership and harmonic stability play an important role in the perception of similarity based on harmony, as well as the common pitch classes shared by two chords. Every correlation of measurements on chord level was higher than those using root notes only, which implies that more complex chord representations can enhance the performance of harmonic similarity measurements. A combination of results from the two highest-ranking models, hierarchy of harmonic stability and TPS, provides even higher correlation with the human judgements. This suggests that considering multiple tonality models can yield to better results than relying on a single model only.

IV. CONCLUSIONS

We presented five different types of chord sequence similarity measures based on different psychological distances of harmony. We also gathered human judgements of similarity between two chord sequence pairs. The main contribution of this study is to provide theoretical and empirical evidence on the relationship between computational approaches and human judgements on harmonic similarity. Furthermore, the results show that using a combination of similarity measures from different tonality models can improve correlation with human survey data. However, there are still numerous issues remaining to be addressed for the measurement of chord sequence similarity. In this study, we handled the substitution cost only while fixing the deletion and insertion costs to a constant value. Considering all cases separately and in conjunction will help comparing more diverse chord sequences and the importance of each operation. Finally, future work needs to take contextual information within a chord sequence into account.

REFERENCES

Bharucha, J., & Krumhansl, C. L. (1983). The representation of harmonic structure in music: Hierarchies of stability as a function of context. *Cognition, 13*(1), 63-102.

Casey, M. A., Veltkamp, R., Goto, M., Leman, M., Rhodes, C., & Slaney, M. (2008). Content-based music information retrieval: Current directions and future challenges. *Proceedings of the IEEE, 96*(4), 668-696.

Cook, P. R. (1999). *Music, cognition, and computerized sound: An introduction to psychoacoustics*: The MIT press.

De Haas, B., Veltkamp, R. C., & Wiering, F. (2008). *Tonal Pitch Step Distance: a Similarity Measure for Chord Progressions.* Paper presented at the ISMIR.

De Haas, W. B., Wiering, F., & Veltkamp, R. C. (2013). A geometrical distance measure for determining the similarity of musical harmony. *International Journal of Multimedia Information Retrieval, 2*(3), 189-202.

Freedman, D. (2015). *Correlational Harmonic Metrics: Bridging Computational and Human Notions of Musical Harmony.*

Hanna, P., Robine, M., & Rocher, T. (2009). *An alignment based system for chord sequence retrieval.* Paper presented at the Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries.

Lerdahl, F. (2001). *Tonal pitch space*: Oxford University Press.

Nordmark, J., & Fahlén, L. E. (1988). Beat theories of musical consonance. *KTH Royal Institute of Technology STL-QPSR, 29*(1), 111-122.

Rocher, T., Robine, M., Hanna, P., & Desainte-Catherine, M. (2010). *A survey of chord distances with comparison for chord analysis*: Ann Arbor, MI: Michigan Publishing, University of Michigan Library.

Film, television, and music: Embodiment, neurophysiology, perception, and cognition

Siu-Lan Tan[1], Mark Shevy[2]
[1]Psychology, Kalamazoo College, Kalamazoo, Michigan, USA
[2]Communication and Performance Studies, Northern Michigan University, Marquette, Michigan, USA

With the rapid, global growth of audio-visual media use in daily life, theory and research focusing on the perception and cognition of multimedia presentation has never been more important or more timely. Until recently, however, there has been little empirical focus on the role of music in the multimedia experience, though the work in this area is gaining momentum (Tan, Cohen, Lipscomb, & Kendall, 2013). The research presented in this symposium breaks new ground in theory, methodology, and empirical investigation with respect to the unique physiological and mental processes involved when humans engage with music and moving images in the context of film and television. The first presentation applies analytical frameworks of embodied cognition to various facets of the temporal domain of music within film. Departing from the behavioral measures typically used in the study of music and multimedia, the next presentation provides data collected using magnetoencephalography (MEG) to explore how the combination of audio and visuals in film may stimulate brain activity that differs from that caused by audio or visuals alone. Delving further into the semantic resonances of the film-music relationship, the third presentation considers film-music pairings, with particular attention to pairs perceived as incongruent. The last presentation investigates the possible links between music used in television commercials and various sociodemographic variables. Following the four presentations, the discussant will lead an extended discussion with the presenters and attendees, highlighting the diversity of methods and interdisciplinary character of multimodal research (as represented by our symposium). The discussion will provide the opportunity to raise questions, delve into topics in more depth, and draw rich thematic connections between the speakers' topics and within the field of music psychology at large.

Temporality and embodiment in film music

Juan Chattah[1]
[1]Music Theory and Composition, University of Miami, Coral Gables, Florida, USA

Although most film-scoring techniques have gradually emerged through the intuitive use of music, recent research in embodied cognition lends itself to examining these techniques from theoretical and empirical perspectives: *grounded cognition* (e.g. Barsalou, 2008) maintains that meaning is generated and stored in sensorimotor experiences; *ecological* and *evolutionary* approaches (e.g. Clarke, 2005) take notion of 'affordances' (Gibson, 1983) as a point of departure; *conceptual metaphor* and *schema* theories (e.g. Lakoff & Johnson, 2003) propose that cognitive representations and operations emerge from sensorimotor processes; and the discovery of the *mirror neuron system* in macaques (e.g. Rizzolatti, Fogassi, & Gallese, 2001) reveals neurophysiological mechanisms that underpin social cognition. This presentation focuses exclusively on the temporal domain of music within film, and traces the logic that motivates embodied meaning while shedding light on a wide range of phenomena, including parallels between the visual and aural domains, the effect of tempo fluctuations in arousal level, the role of beat anticipation and predictability in relation to perceptual judgment of closure and stability, and the influence of meter in depicting collective motor coordination that may lead to social bonding and the construction of individual and group identities. By applying analytical frameworks from embodied cognition to a number of film music examples, this presentation extends beyond cognitive mappings of physicality to include cultural practices and values, and elucidates how musical temporality supports, highlights, or comments on other facets of the cinematic experience.

Neurophysiological responses to motion pictures: Sound, image, and audio-visual integration

Roger E. Dumas[1,2], Scott D. Lipscomb[2,3], Arthur C. Leuthold[4], Apostolos P. Georgopoulos[1,2,4]
[1]Brain Sciences Center, VAMC, Minneapolis, Minnesota, USA
[2]Center for Cognitive Sciences, University of Minnesota, Minneapolis, Minnesota, USA
[3]School of Music, University of Minnesota, Minneapolis, Minnesota, USA
[4]Neuroscience, University of Minnesota, Minneapolis, Minnesota, USA

Past investigations into the influence of a musical soundtrack on audience cognition of movies, animation, video games, and other multimedia have focused primarily on behavioral measures of the experience based on verbal responses (e.g., semantic differential scales). It is important to supplement such cognitive-perceptual research with studies that inform us about the neurophysiological activity present during the viewing of motion pictures. In the present study, magnetoencephalography (MEG) was used to examine the whole brain in identifying significant relations between dynamic, interacting neural networks and systematically varied audio-visual stimuli, namely five audio-visual excerpts from the feature film *Star Trek IV: The Voyage Home* in each of three conditions: (a) musical soundtrack-only, (b) visual-only, and (c) music and visual combined. All data have been collected and are currently being analyzed. Upon completion of our analysis, we expect to identify regions of the brain that were not activated in the audio-only or visual-only conditions, but do become activated in the audio-visual condition. Completion of the ongoing study will facilitate our understanding about the various areas of the brain that are activated differentially when comparing neurophysiological activity in audio only, visual only, and audio-visual conditions. These results will further inform our knowledge regarding the role of music in the motion picture experience, augmenting our growing understanding related to the behavioral-cognitive responses that have been much more thoroughly investigated in past research studies investigating multimedia cognition.

Exploring incongruence: Shared semantic properties and judgments of appropriateness in film-music pairings

David Ireland[1] and W. Luke Windsor[1]
[1]School of Music, University of Leeds, Leeds, UK

Ideas of congruence recur in studies that address the impact of various properties of the film-music relationship on audience perception and response, and relate to shared properties of structural, semantic or holistic dimensions of the stimuli. Less research has focused on audiovisual incongruence, which is often conflated with judgments of inappropriateness in a film-music relationship. Yet, such judgments are multi-dimensional, context-dependent and subjective. This study explores the relationship between the perceived appropriateness of film-music pairings, and the number of semantic properties shared by the component extracts that comprise them. Participants (N=40) rated four music and four film excerpts, which reflected differing states on visual analogue scales labeled "agitated-calm" and "happy-sad." They also rated all possible combinations of these excerpts (i.e., 16 pairings), using the same scales and a third scale labeled "audiovisual pairing is appropriate-audiovisual pairing is inappropriate." Results show that, in general, pairings with shared properties on both the "agitated-calm" and "happy-sad" scales were rated as more appropriate; pairings with fewer shared properties were rated as less appropriate. Significant differences exist between appropriateness ratings for pairings that share properties on both "agitated-calm" and "happy-sad" scales and pairings that share properties on just one of these scales, or on neither scale. Yet, some of the film-music pairings challenge these broader patterns, including those using film rated as "calm." Focusing on one such pairing, which combines film from *The Shawshank Redemption* and music from the score for *Gladiator*, this paper will question ideas of additivity used to explain film-music interactions. The researchers will contend that a psycho-semiotic approach, involving musicological, semiotic and conceptual analyses can complement empirical studies to more holistically account for the complexity of film-music incongruence.

Classical music in television commercials: A social-psychological perspective

Peter Kupfer[1]
[1]Department of Music History, Southern Methodist University, Dallas, TX, USA

Previous studies have uncovered relationships between listeners' musical preferences and certain sociodemographic factors (e.g., sex, age, income, education, and musical training). Similarly, much of television advertising is targeted at consumers based on relevant sociodemographic categories. This study thus aimed to measure differences between sociodemographic groups regarding their reception of commercials (as measured by appeal and congruency ratings) based on the type of music employed. Because classical music is a relatively clearly defined genre that is recognizably different from most popular styles, and because, as previous studies have shown, its listeners typically fall into clear sociodemographic groups (e.g., older, richer, and more educated), the focus of this study was on classical music, though distractor tracks using various other styles were included as well. The 557 respondents were divided into four sociodemographically equivalent groups as follows: Group 1 listened only to the musical tracks of the stimuli (i.e., without image tracks); Group 2 watched only the video tracks (i.e., without music); Group 3 watched the original commercials with music and video; and Group 4 watched "recombined" commercials in which music tracks were swapped. The stimuli comprised nine nationally aired commercials for products ostensibly appealing to different sociodemographic groups. In order to control for the effect of music, the chosen commercials consisted largely or only of music and images, five using classical music, four using other styles. The results of the study showed that while there were significant correlations between respondents' sociodemographic background and their preferences for certain musical genres, and between their sociodemographic background and preferences for certain ads/products, these differences disappeared when music and images were combined, regardless of the musical style used. The findings suggest that sociodemographic differences do not play a central role in the processing of music in commercials, classical or otherwise.

The Role of Idiomaticism and Affordances in Bebop Improvisation

David John Baker,*[1] Adam Rosado,*[2] Eva Shanahan#[3] Daniel Shanahan*[4]

*School of Music., Louisiana State University, USA

#Independent Researcher., Baton Rouge, USA

[1]dbake29@lsu.edu, [2]arosad2@lsu.edu, [3]shanahan.eva@gmail.com, [3]dshanahan@lsu.edu

ABSTRACT

Improvisational styles rely on an implicit knowledge of various musical gesture, ranging from not-to-note transitions to those that exist on a more structural level. These prototypical musical gesture are linked and transformed over time, and are intertwined with instrumental idiomaticisms, which are derived from the structure of the instrument and the amount of effort required by certain musical gestures (see Huron and Berec, 2009; Gjerdingen, 2009). This study explores how these stylistic and physical constraints interact in bebop improvisation –specifically the transcribed solos of Charlie Parker, Dizzy Gillespie, an Clifford Brown. The aims of this study are fourfold. Firstly, we aim to replicate the finding of Huron and Berec (2009), which found that keys chosen for compositions correspond with the most idiomatic transitions. Secondly, we hypothesis that an "arc of idiomaticism" occurs throughout the course of an improvisation. Thirdly this study examines the most common structural schemata in the corpus, looking for the most common occurrences of temporally regular gesture, which were then categorized to create a schema-derived taxonomy. Lastly, this study examines the interaction of the idiomatic gestures with the most common improvisatory schemata in the corpus. Examining relationships between stylistic usage and affordant idiomaticism. This study used data taken from trumpets and saxophonists to provide a metric for the relative difficulty of not-to-note transitions, taken from the corpus. A web-based study then asked self-identified trumpeters and saxophonists to describe the relative difficulty of certain note-to-note transitions. A difficulty metric was then attached to each note-to-note transition in a corpus of bebop improvisations encoded in *kern format (Huron, 1995). This study suggests that use of improvisational corpora can serve as a novel medium to further explore the relationship between stylistic and physical constraints.

I. INTRODUCTION

Theories of jazz improvisation often fall into two broad camps: ones of schema-derived gestures ("licks", motives, etc.), or ones of motor associations with sound that facilitate the creation of such gestures. Not only are these two not mutually exclusive, but they are also quite interdependent. For example, a schema might first become widely used because of physical, motor gestures that allow for the pattern to be played easily by a wide range of musicians. The current study examines the the interaction between these two improvisational aspects in bebop improvisation, but specifically focuses on the role of instrumental affordances.

Owens (1974) examined the role of patterns in bebop (specifically Charlie Parker improvisations), drawing links to the tonal structure of the piece, but also creating a network between these gestures throughout the entire Charlie Parker output. Johnson-Laird (2002) argues that these patterns are stylistic constraints governed by the underlying harmonies. He constructs a computational model of improvisations that focuses primarily on the immediate transitions of notes and their relationship to an immediate harmony, forgoing large scale patterns. (2002). Pressing (1988) suggests that it is a more momentary event, in which the production of an idea is influenced by those immediately preceding it. Pressing also speaks of the "underlying formal scheme" of a piece, as well as the motives, arpeggios, and patterns needed to perform it (Pressing, 1984). Norgaard's (2014) work on Charlie Parker extends the idea of note-to-note transition probabilities and included specific intervallic patterns performed by bebop improvisers. Specifically, Norgaard found that a vast majority of Parker's improvisatory gestures begin a four-interval pattern, and that a limited set of interval and rhythmic relations make up a large part of his improvisations.

The current study focuses primarily on this aspect of note-to-note transitions, and investigates the role of instrumental affordances on stylistic choices. We first aim to extend some of the work on the physical constraints of an instrument by attempting to obtain difficulty ratings of note-to-note transitions in a different, yet comparable way (see Huron and Berec, 2009). We similarly present the findings of a study in which participants were asked how idiomatic specific note-to-note transitions were on their instrument, and then fit those ratings onto encoded corpora of jazz improvisations, namely Charlie Parker, Dizzy Gillespie, and Clifford Brown. A more robust rating system would then hopefully be able to be applied to any sort of corpora to obtain difficulty and further explore instrumental idiomaticisms. The following paper is divided into two sections following our four main hypothesis listed below. We hypothesize that:

- Keys chosen for composition and improvisation will correspond to the most idiomatic or least difficult keys.
- Given the difficulty of transition ratings, we will find an "arc of idiomaticism" where improvisers will make musical choices more specific to their instrument throughout a solo?

Future work will use this data to address the questions pertaining to more top-down structures, such as schemata, and will specifically address the following hypotheses:

- Given melodic, bass, and harmonic information we will begin to recognize common schematic patterns used in improvised solos of Charlie Parker and Clifford Brown.
- Stylistic choices made between soloist given certain schemata will be objectively more difficult when mapped on to affordant ratings of other instruments.

II. CORPORA USED

This paper also introduces two new corpora for the computational musicology community. The first is the Charlie Parker Omnibook (Aebersold and Slone, 1978). While not entirely new as a melodic corpus, having been used in previous studies before, this corpus contains solos (N=59) with added in harmonic information and metadata (such as recording information, and information on the sidemen) in *kern format for processing using David Huron's Humdrum toolkit (Huron, 1995). We also introduce solos (N=57) from the Complete Transcriptions of Clifford Brown by Marc Lewis (1991). While equal in number of solos compared to the Omnibook, this Brown corpus contains almost twice as much data and additionally contains realized chord progressions. We also use some transcriptions of Dizzy Gillespie (N=12). All corpus material is available for use upon request.

III. IDIOMATICISM AND KEY CHOICE

Our first task was to replicate the findings from Huron and Berec (2009) that showed that compositional choices were found to be more idiomatic for the instrument if the composer of the piece were themselves considered a composer for that instrument. Our study is not an exact replication of this in that we rely on improvisatory material as opposed to composed music. Below we outline how this data was obtained. These ratings are then used throughout the rest of the studies.

A. Obtaining Ratings of Idiomaticism

A web study was set up to collect data on perceived difficulty of note-to-note transitions by instrument from advanced players. Respondents who identified as advanced level players on alto saxophone (N=11), tenor saxophone (N= 7) and trumpet (N=15) made ratings on random subsets of a collection of notes presented on the treble clef from a pool of every note transition from G3 to G5. *A priori* we decided to only use natural, sharp, and flat notes, excluding all double sharps and flats since this study was not based on orthography. Participants were presented with the musical notation and asked to make a rating on the keyboard using a 7 point Likert scale with 1 indicating "One of the Easiest Intervals to Play" and 7 indicating "One of the Most Difficult Intervals to Play". Participants also had the option to respond with a 0 to indicate that an interval was unplayable on their instrument since the pool of notes used was not specific to either instruments range and will serve as the basis for collecting future data on note-to-note transitions on other instruments.

Ratings from each instrument were collected, pooled, averaged, and then mapped on to each corpus. Solos from each corpus were additionally transposed up and down 11 half steps with the appropriate difficulty ratings being mapped on to the new note transitions. Solos were transposed both up and down to further explore effects of range.

B. Improvisation and Key Choice

First, we examined the works of a trumpeter (much like the Huron and Berec study), performing a t-test between the original key Clifford Brown played his solos and two half steps up and down comparing average idiomaticism ratings of the original key with its transpositions. On average solos played at Brown's original transposition scored an average lower with a rating of 1.31 compared to the 1.39 rating from a half step above ($p<.001$). Interestingly, no statistically significant difference was found in the idiomaticism ratings when transposed a half step above. This could be due to the fact that the study took all pieces and averaged them together, or we simply need more musicians to rate the pieces to provide a more accurate model.

We decided to also examine the relationship between the instruments transposed pitch and its concert pitch. We performed a t-test between the original key Charlie Parker played his solos and two half steps up and down comparing average idiomaticism ratings of the original key with its transpositions. On average original key solos scored an idiomaticism ratings of 1.25 with a half step above scoring on average 1.48 ($p < .001$) and a half step below scoring 1.47 ($p<.001$). Significance was still maintained after using the Bonferroni correction for multiple t-tests.

Additionally difficulty ratings for each instrument were mapped onto each corpus to explore the effect of difficult ratings on transpositions. The first three charts below (Figures 1-2) show how ratings from the three instruments map on to the Charlie Parker Corpus. Although individual differences between keys are not overwhelming, there is a clear effect of range.

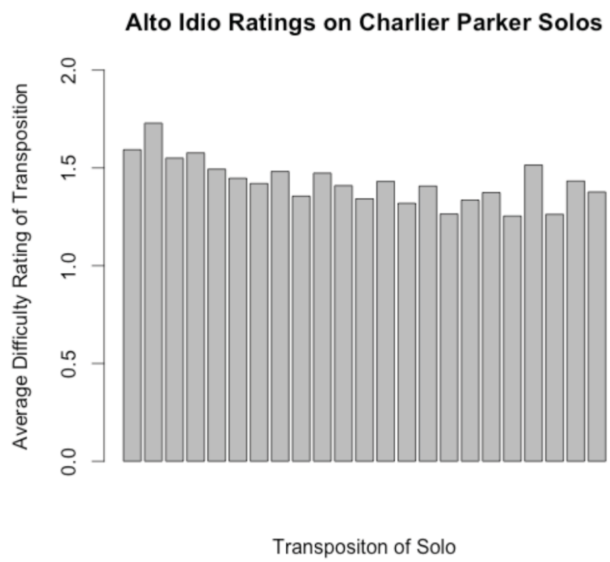


Figure 1. Difficulty ratings mapped onto Charlie Parker Solos. The middle of the graph is the untransposed rating, with transpositions up twelve semitones to the right, and down twelve semitones to the left.

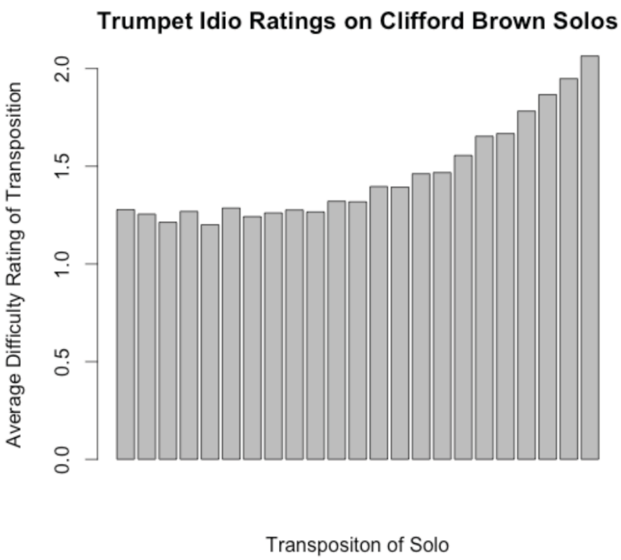


Figure 2. Trumpet Ratings Mapped to Clifford Brown Corpus. Difficulty ratings mapped onto Charlie Parker Solos. The middle of the graph is the untransposed rating, with transpositions up twelve semitones to the right, and down twelve semitones to the left.

C. Discussion

This first study yielded interesting results, but we would ideally have more data from subjects. As it currently stands, simply due to the sheer number of intervals presented to participants, very few intervals were rated by more than 3-4 players. We would ideally have enough data for each note transition to have a more reliable estimate. While the results support the notion that transpositions on the saxophone were less idiomatic when you moved them away from the original key, we had difficulties with the trumpet data. There is obviously an effect of instrumental range (especially for the trumpet); as things are mapped to the higher register difficulty ratings increase, an effect that we do observe with the saxophone difficulty ratings (see Figure 2).

IV. ARC OF IDIOMATICSIMS

Our second hypothesis argued that there would be a trajectory of idiomaticism throughout an improvisation: pieces might exhibit an arc, or perhaps simply a linear relationship over the course of an improvisation. We figured that soloists would play more difficult intervals as the solo would go on, perhaps as a way of generating more interest during repeated choruses. In order to look at this, we I took solos and examined their placement over the course of the improvisation with the Humdrum Toolkit (Huron, 1995). As jazz improvisations (like most music) consists mainly of step motion we discarded the ratings that were a "level 1", as we were concerned with getting a floor effect.

As can be seen from Figure 3 (the solo to "52nd Street"), we originally had reason to believe that a linear relationship might be present, although the trend was not significant. Interestingly, a decreasing (but similarly non-significant) trend was found over the course of Gillespie's "Caprice" solo.

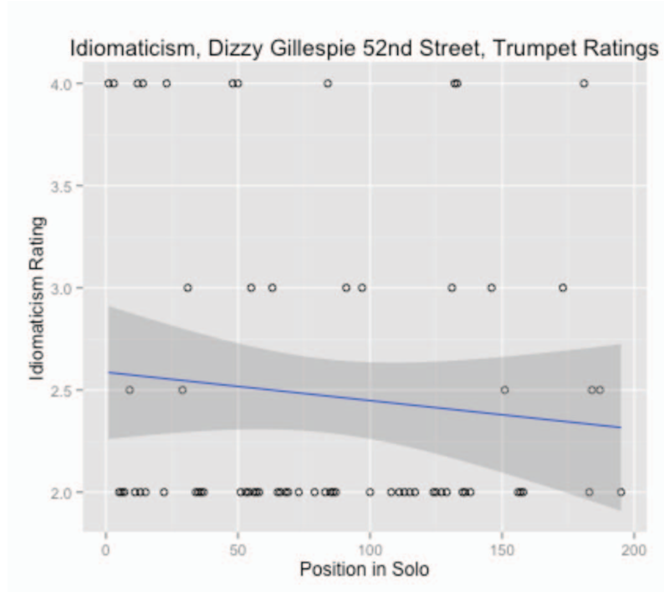


Figure 3. Difficulty ratings of note-to-note transitions (in the original key) over the course of Gillespie's "52nd Street" solo.

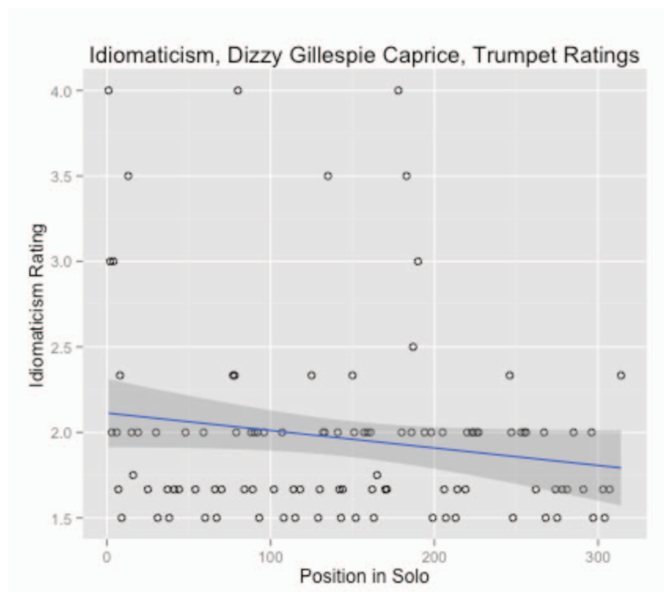


Figure 4. Difficulty ratings of note-to-note transitions (in the original key) over the course of Gillespie's "Caprice" solo.

Surprisingly, the ratings do often exhibit a downward trend (contrary to our hypothesis) when mapped onto a mismatching instrument. For example, in Figure 5, we can see the difficulty ratings decreasing over the course of Gillespie's "Salt Peanuts" solo. This effect is not replicated throughout the corpus, and it's difficult to find any such consistent effects.

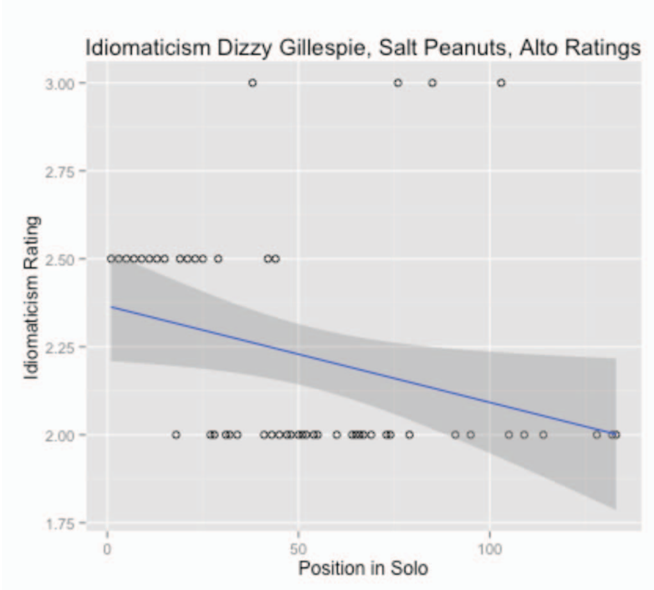


Figure 5. Alto saxophone difficulty ratings of note-to-note transitions (in the original key) mapped onto Gillespie's trumpet solo over the course of "Salt Peanuts".

V. CONCLUSION

Our current study attempted to demonstrate the role of note-to-note affordances on improvisation. While there does seem to be some effect on key-choice, it is not always replicable when going up a half step when looking at trumpet ratings, a result that came as a surprise to us (including the jazz trumpeter in the group). Our second hypothesis regarding a trend in difficulty ratings over the course of an improvisation showed no significant trends in either direction, and we therefore failed to reject our null hypothesis. Future work will gather more data on these note-to-note transitions, in the hopes of having more reliable ratings, and will expand to look at the interactions between difficulty ratings, entropy measures, and rhythmic aspects, as well as the top-down effects of harmony, improvisational interaction, and schema.

ACKNOWLEDGMENT

We would like to thank Craig Sapp for his assistance with stimuli design and a transposition script.

REFERENCES

Aebersold, J., & Slone, K. (1978). Charlie Parker Omnibook. *Atlantic Music Corp*.

Gjerdingen, R.O. (1986). The formation and deformation of classic/romantic phrase schemata: A theoretical model and historical study. *Music Theory Spectrum, 8,* 25-43.

Goldsen, M. H. (1978). Charlie Parker Omnibook: For C Instruments. *Atlantic Music Corp.*

Huron, D., & Berec, J. (2009). Characterizing idiomatic organization in music: A theory and case study of musical affordances.

Huron D. B. (1994). *The Humdrum Toolkit: Reference Manual.* Center for Computer Assisted Research in the Humanities.

Johnson-Laird, P. N. (2002). How Jazz Musicians Improvise. *Music Perception: An Interdisciplinary Journal*, 19(3), 415-442.

Lewis, M. (1991). Clifford Brown Complete Transcriptions. *Chas Colin.*

Norgaard, M. (2014). How Jazz Musicians Improvise: The Central Role of Auditory and Motor Patterns. *Music Perception: An Interdisciplinary Journal, 31(3),* 271-287.

Owens, T. (1974). *Charlie Parker: Techniques of Improvisation* (Vol. 1). University of California at Los Angles.

Paparelli, F. (1975). Dizzy Gillespie, A Jazz Master. *Hal Leonard Corporation.*

Pressing, J. (1984). Cognitive Processes in the Perception of Art. W.R. Crozier and A.J. Chapman (editors). *Cognitive Processes in the Perception of Art*, *19*, 345.

Pressing, J. (1988). Improvisation: methods and models. *John A. Sloboda* (*H.g.*): *Generative Processes in Music, Oxford*, 129-178.

Improvisation as a way of knowing

Andrew Goldman[1]
[1]Presidential Scholar in Society and Neuroscience, Columbia University, New York, NY, USA

The lack of a single widely accepted definition of improvisation has not hindered the development of the field of improvisation studies. In fact, a critical literature is flourishing by questioning its historical, sociological, and philosophical foundations. However, in order to do scientific experiments on improvisation, an interim concrete definition must be chosen. Much psychological and neuroscientific work has defined the phenomenon in the problematic and culturally contingent terms of novelty and spontaneity. In this theoretical paper, I argue that without being more critically sensitive to concepts like "novel" and "spontaneous," the scientific theories that are built around them remain equally ambiguous and problematic. I propose a different conceptual foundation to guide future empirical work. Improvisation can be considered as a way of knowing. Cognitive-scientific research has elsewhere shown that musical training changes how people perceive and perform music (i.e., how they know about it). Given that musicians learn through different methods, it is reasonable to theorize that there are differences between the ways musicians know about and use musical structures. Improvisers train in ways that emphasize links between perception and action, not just in expressive aspects of performance, but also learning which movements create which harmonies and melodies in a way that may differ from score-dependent musicians. Improvisers may have different knowledge of sensorimotor contingencies. Perception-action coupling experimental methods could be adapted to compare groups of improvisers and "non-improvisers," categorized on the basis of questionnaires that evaluate practice methods and musical experiences. Exploring these cognitive differences can serve as a productive alternative foundation for empirical work that characterizes what it means to be an improviser, and how improvisers acquire and use their knowledge.

Perception of Structural Features in First and Second Musical Themes

Lindsay Warrenburg,[*1] David Huron,[*2]

[*]School of Music, Ohio State University, USA

[1]warrenburg.3@osu.edu, [2]huron.1@osu.edu

ABSTRACT

In music exhibiting two main themes, the first theme is commonly characterized as stronger or more energetic, whereas the second theme is characterized as gentler or more cantabile. Previous work indicates that compared to first themes, second themes are less likely to be in the minor mode, but are more likely to be legato, make use of a slower tempo, utilize smaller intervals, and involve a quieter dynamic level. The current study tested whether musicians are sensitive to these changes in the musical structure. Musicians (n = 44) were tested on their ability to discriminate between pairs of musical themes in 22 piano works. Participants completed two tasks, one based on the musical score and one based on sound recordings. The study was designed as a two-alternative forced-choice task, where musicians were asked to identify which of the two themes comes first in the music. Additionally, the type of information available in the musical excerpts was manipulated so that in some trials, participants did not have access to dynamic and articulative information. The results indicate that musicians are weakly able to differentiate between first and second themes in both the score-only (p < 0.001) and audio-only tasks (p < 0.001), and that participants perform better when stimuli include dynamic and articulative information (p < 0.02). The results are consistent with the claim that musicians can weakly perceive structural and expressive differences between first and second themes, even outside of the original musical context.

I. INTRODUCTION

Contrasts between first and second musical themes have been described in musicological and music theory writings over the past 300 years. Theorists in the eighteenth and nineteenth centuries often described contrasting musical themes using descriptive and metaphorical terms. Common descriptions of first themes included terms like *strong* and *energetic,* whereas typical descriptions of second themes included terms such as *lyrical* and *gentle* (Churgin, 1968; Koch, 1793/1983; Kollmann, 1799; Newman, 1963). Modern theorists tend to emphasize how the structure of themes can differ from one another (Caplin, 1998; Hepokoski & Darcy, 2006; Rosen, 1971). A corpus study conducted by Warrenburg and Huron (2015) aimed to empirically test whether these common theoretical ideas regarding first and second themes could be observed in a large sample of music. The excerpts of music tested included a broad range of musical forms, genres, and ensembles. The empirical findings of Warrenburg and Huron (2015) were consistent with the hypothesis that, compared to first themes, second themes are less likely to be in the minor mode and are more likely to be legato, make use of a slower pace, use smaller intervals, and utilize a quieter dynamic level.

In light of the apparently stable features distinguishing first and second musical themes, it is of interest to test whether these features are perceptually salient for listeners. The current study aimed to determine whether people can differentiate between a pair of musical themes using both visual and audible factors. The study focused on the relationship between first and second themes from the same musical work and consisted of music composed in the eighteenth through twentieth centuries.

II. HYPOTHESES

The question at hand is to determine whether first and second musical themes can be perceptually distinguished. In daily life, this task most directly applies to how listeners hear musical performances. However, the way in which a performer brings a piece of music to life can exaggerate, eliminate, or alter the musical criteria written in the score. For example, existing knowledge of sonata form could lead a performer to exaggerate the *energetic* first themes and the *lyrical* second themes. Alternatively, performers may wish to convey information that they believe is lost in the musical score, based on extra-musical knowledge. They may also wish to provide a new interpretation of a well-known piece. Playing a typically *graceful* second theme in a sardonic manner will obscure the musical content provided in the score. Therefore, any hypothesis must be made in regard to music heard without interpretive performance nuance. This leads to the first hypothesis:

> *H1.* Participants can categorize first and second themes when listening to MIDI sound recordings.

Un-interpreted MIDI recordings take score-based data and produce an auditory recording without any performance nuances. However, with the advent of recent technology, it is possible to have MIDI recordings that differ in features like tempo, dynamics, and articulation. For the purposes of this study, we were mostly interested in how variations in articulation and dynamics affected performance on the discrimination task. Since the inclusion of articulation and dynamics is likely to help in the discrimination between first and second themes, a corollary to the first hypothesis was made:

> *H1b.* People are better at categorizing themes from sound recordings when dynamics and articulation are included than from sound recordings without these features.

An even more reductive approach can be taken when considering the written musical score. Discriminating between first and second themes from a score requires a different set of skills than does an aural analysis. However, the notated music likely does contain information that can allow a person to distinguish between two written themes. The interest in musical scores leads to a second hypothesis and its corollary:

> *H2*. Participants are able to distinguish first and second musical themes by looking at the information contained in the musical score.
>
> *H2b.* People should perform better on the task when the provided scores include dynamics and articulation than when the given scores do not include this information.

Since the analysis of written scores requires familiarity with musical notation, and therefore some training, it could be expected that people with more musical training would do

ISBN 1-876346-65-5 © ICMPC14

considerably better than people with less musical training in any score-based perception task. Although musical training should also facilitate performance on the recording-based perception task, the difference in performance between people of various levels of musical training may be smaller in the case of the recording-based task. No hypothesis was made about whether people with a high level of musical training would be better at the score-based or recording-based perception tasks, since, to some degree, they are different skills.

As mentioned in the introduction, the structural features operationalized and measured in Warrenburg and Huron (2015) have been shown to differ between first and second themes. These features included durational pace (a proxy for tempo), average interval size, rhythmic smoothness (measured using the nPVI), mode, key, articulation, and dynamics. Accordingly, when differentiating between two themes, musicians should be able to use these criteria to make their judgments. Additional criteria may be used to make these distinctions, as well, but no *a priori* hypotheses were made regarding other criteria. Importantly, the consideration of harmony and phrase-structure were not considered as *a priori* hypotheses. In order to compare the perceptual and corpus studies, it was decided to retain the same criteria in the perceptual study as was tested in the preceding corpus study. Harmonic implications will be discussed more at the end of the paper.

III. SAMPLE

For the purposes of this study, a sample of 22 piano works was taken from Sam Barlow and Harold Morgenstern's *Dictionary of Musical Themes* (1948). Sampled items were limited to works containing precisely two themes. That is, pieces containing an introductory theme or entailing more than two themes were excluded. Approximately half of the sample consisted of works in sonata form, with the remaining half written in other structural forms. Sampled piano works were composed between 1777 to 1939, with most works composed during the Classical and Romantic periods. Using the Humdrum Toolkit (Huron, 1993), the normalized pairwise variability index (Low, Grabe, & Nolan, 2000), average pace value, average interval size, mode, and key were estimated. In addition to these measures, manual coding of articulation and dynamics was carried out for all 22 piano works. Dynamics were coded using the traditional Italian levels: *ppp, pp, p, mp, mf, f, ff,* and *fff. Sforzandos,* crescendos and diminuendos were ignored. Only the basic dynamic level was coded. In the analysis, the eight levels (identified above) were treated as ordinal data. That is, *ppp* was coded as 1, *pp* as 2, *fff* as 8, etc. Similarly, articulation markings in the thematic passages were subjectively coded as one of five possible designations, again coded using an ordinal scale: very legato (coded 1), generally legato (coded 2), balanced/unclear (coded 3), generally staccato (coded 4), and very staccato (coded 5).

The comparison of thematic pairs from a single musical work could lead musically-trained participants to respond in one of a few ways. The first outcome is that participants could distinguish between the musical cues written by the composer, using one or more characteristics of the music in order to make their judgment. Participants may use subtle criteria, such as variations in interval sizes, articulation, and dynamics to distinguish between two themes. Another possibility is that participants use a heuristic to categorize first and second themes. A common heuristic would be that a major theme is a second theme, whereas its paired minor theme is a first theme, given conventions about primary and secondary themes in sonata form. Attempts were made to circumvent the use of such heuristics. Therefore, the sample consisted of first and second themes that are in the same mode. Additionally, all of the samples were transposed to have a tonal center of C: musical pairs that were in a major mode were transposed to C major and musical pairs that were in a minor mode were transposed to C minor. The transposed samples prevented key relationships (such as I – V or i – III) from affecting the discrimination between first and second themes. Therefore, the categorization of a pair of musical themes must be done with regard to more subtle musical criteria, instead of only using basic knowledge about music theory.

IV. METHOD

The study consisted of two parts, both based on a two-alternative forced-choice task. The first part asked participants to examine the musical score of a pair of musical themes that came from the same work; they were asked to identify which theme they believed comes first in the music. The second part requested the same participants to listen to recordings of the same musical thematic pairs and to identify which theme was presented first in the music. Both the score-based task and the recording-based task consisted of two conditions: A) musical samples with dynamic and articulative information and B) musical samples without dynamic and articulative information.

A. Score-Based Perception

One of the hypotheses of this study is that participants are able to categorize first and second musical themes by looking at the information contained in the musical score. The task used to test this hypothesis required participants to distinguish between a pair of themes that came from the same musical work. In order to avoid variability between different published musical editions, all themes were reprinted using the same notation software (Finale 2014). The themes were notated as they first appear in the musical score. All external cues were eliminated from the Finale notation. This included types of information such as measure numbers, tempo markings, and headers like "Minuet" or "Trio." Performance indications such as "animato," "dolce," and "furioso" were also removed from the thematic samples. In this study, accompaniment of the theme was included, allowing some harmonic and textural cues to be available to the participants. Further discussion of this methodological choice is deferred to the end of the paper.

The hypothesis also specified that participants should perform better on this task when they are given access to articulation and dynamic information. Therefore, two versions of the musical samples were notated using Finale 2014. The first version consisted of the musical notes, dynamic markings, and articulation markings; this version was deemed the "With Dynamics and Articulation (With DA)" sample. A second version removed the dynamic and articulation markings and was deemed the "Without Dynamics and Articulation (Without DA)" sample. Therefore, in the "Without DA" sample, participants could only use certain features of the musical structure (such as nPVI, average interval size, and pace value) to make the determination of "first" and "second" themes. An example of each of the two conditions is shown in Figure 1.

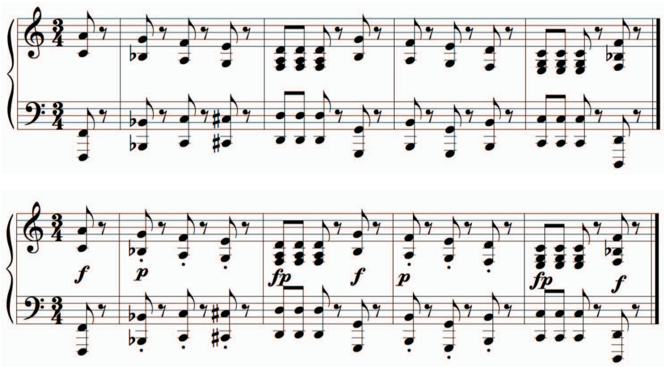



Figure 1. The same musical score shown in the "Without Dynamics and Articulation" condition (top line) and in the "With Dynamics and Articulation" condition (bottom line). The music is the first theme of Dvořák's *Waltzes, Op. 54, No. 3.*

Participants were presented with two themes from a musical work on a single screen. They were asked to categorize the pair of themes by indicating which theme they predicted came first in the music and which theme they judged to appear second in the music. Additionally, they were asked to rate the confidence of their response on a 10-point scale. The order of presentation for the musical works was randomized, as was the order of the first and second themes for each musical work.

In an attempt to circumvent demand characteristics, it was decided that each participant would not be presented with both the "With DA" and "Without DA" versions of each musical pair. Instead, the study consisted of two experimental groups, utilizing a between-subjects design. The groups were counterbalanced with regard to the dynamic and articulative condition. In Participant Group 1, 11 of the musical pairs were given "With DA" and 11 musical pairs were shown "Without DA." In Participant Group 2, these conditions were reversed, so that the musical pairs that were in the "With DA" condition in Participant Group 1 were the musical pairs that were in the "Without DA" condition in Participant Group 2. Therefore, every musical thematic pair was used in the experiment with dynamics and articulation and without dynamics and articulation, but each participant only saw each thematic pair in one of the two conditions. This design allows overall performance of the music categorization to be compared across the two conditions. Presumably, people might be expected to perform better in the "With DA" condition than in the "Without DA" condition.

B. Recording-Based Perception

The second major hypothesis is that listeners should be able to discriminate between first and second themes from audio recordings. The corollary hypothesis is that recordings including dynamics and articulation should result in better overall performance than recordings without these two features. Therefore, similar to the score-based perception task, the recording-based task involved two conditions: one with the inclusion of articulation and dynamic information ("With DA") and one without these features ("Without DA").

In order to compare a single participant's performance on the score-based and recording-based perceptual tasks, a within-subjects design was maintained across the two major parts of the study. This means that for a given musical sample, the condition (either "With DA" or "Without DA") was maintained across both parts. If a participant saw the score to a Mozart sonata *with* articulation and dynamics and saw the score to a Beethoven sonata *without* articulation and dynamics, he or she would hear the recording of the Mozart sonata *with* articulation and dynamics and hear the Beethoven recording *without* articulation and dynamics. The musical recordings were therefore counterbalanced in a parallel manner to the presentation of the notated scores. In summary, every musical recording in the experiment was heard with and without dynamics and articulation, but each participant only heard one of the two versions of the recordings.

To eliminate bias from real human performance, each recording of the piano music was made from the Finale notation file, using an automated setting called "Human Playback." The tempo chosen was the one indicated by the score or by judgment of the first author (who was familiar with all of the piano pieces in the sample). Two separate recordings of the piano works were obtained from Finale: a control condition (akin to "Without DA") and a performance condition (akin to "With DA"). In the performance condition, the audio file included articulation and dynamic markings.

The design of the recording-based task was the same as the design of the score-based task. Participants heard audio files of the first and second themes from the twenty-two piano works. They were told they could listen to each theme as many times as they wished. Their task was to indicate which theme came first in the music and which theme appeared second in the music. They were also asked to rate how confident they were about their response using a 10-point scale. The order of presentation was randomized, as was the order of the first and second themes for each musical work. Two musical samples were repeated during the course of the experiment in order to collect test-retest reliability.

C. Protocol

Participants completed the two tasks online, via the Qualtrics Research Suite. In order to account for individual variability, the study was created as a within-subjects design, so participants completed the recording- and score-based tasks with the same conditions of the musical pairs (either with or without dynamic and articulative information). Once again, participants were asked to indicate which of the two themes was first and their confidence.

After the participants completed both the score-based and recording-based tasks, listeners were asked to rank-order the possible criteria they may have used in making their determinations. In addition to the musical criteria from the corpus study (excluding mode and key, which were fixed), five blanks were provided so that participants could include other criteria in their rankings. Participants were also asked whether first or second themes were easier to classify (generalizable/had more of a prototype) and whether it was easier to classify first and second themes from the notated scores or from sound recordings. A final question was posed where participants were asked to comment on the relationship between thematic order (first and second themes) and thematic hierarchy (main theme versus subservient theme).

After completing these tasks, participants were asked basic demographic questions about their musical experience – estimated years of music theory training, years of instrumental or vocal training, amount of time per week spent listening to classical music, whether or not they were familiar with sonata-allegro form, and their overall familiarity with Finale and MIDI recordings.

V. RESULTS

A. Demographics

Forty-four participants, primarily from the Ohio State University School of Music, took part in the experiment. The median age was 20 (range from 19 to 39). Of these participants, twenty-one (48%) were female. Participants exhibited a wide range of musical training, with a range of 0-25 years of formal music theory training (mean = 3.66). Six participants had one or fewer years of music theory training. Participants had a range of 1-30 years of formal instrumental or vocal training (mean = 9.43). About half of the participants claimed that they had learned sonata-allegro form (48%) and the majority of the participants had medium or no experience with Finale or MIDI (77% had medium to no Finale experience, 82% had medium to no experience with MIDI).

B. Statistical Analysis

Between the two parts of the study (score-based and recording-based perception), each subject made 44 judgments about first and second themes, with an additional 4 judgments for test-retest trials. With 44 participants, this means that 1936 judgments about first and second themes were made (2112 including test-retest trials). Of these judgments, participants selected the correct answer 57% of the time (1094/1936), which was significantly different from chance, $X^2(1) = 32.802$, $p < 0.00001$. Although the effect size is small, the results are nevertheless consistent with the hypothesis that, on average, participants are able to use structural information from the music to determine which theme from a pair comes first in a musical work.

One of the *a priori* hypotheses was that participants could distinguish first and second themes by looking at the information in the musical score. The data suggest that participants are able to do this at a level statistically different from chance, with 56% of the judgments (539/968) classified as the correct response, $X^2(1) = 12.50$, $p = 0.0004$. The results are shown graphically in the left half of Figure 2. The corollary to this hypothesis was that participants should perform better when the provided scores include dynamic and articulative information than when the given scores do not include this information. The individual differences from both participant groups were combined to use McNemar's test, a nonparametric X^2 test used to compare differences in categorical data. The results were consistent with this corollary hypothesis, with 56.6% correct judgments in the scores with dynamics and articulation and 54.8% correct judgments in the scores without dynamic and articulation markings. Although these values are close in magnitude, they are statistically different using McNemar's test, $X^2(1) = 5.915$, $p = 0.015$. The results are graphically depicted in the left half of Figure 2.

The other main *a priori* hypothesis was that participants could categorize first and second themes when listening to MIDI sound recordings. As shown in the right half of Figure 2, participants were able to perform this task, with 57% of the judgments being the correct answer (555/968), $X^2(1) = 20.831$, $p < 0.0001$. The corollary to this hypothesis was that participants would be better at categorizing themes from sound recordings when dynamics and articulation were included than from sound recordings without this information. The results of this corollary hypothesis were significant according to McNemar's test, with 58.9% correct judgments in the recordings with dynamics and articulation and 55.8% correct judgments in the recordings without dynamics and articulation, $X^2(1) = 9.820$, $p = 0.0017$. The results are shown graphically in the right half of Figure 2.

Participants performed with a slightly higher accuracy in the recording-based task (57.3% correct judgments) than in the score-based task (55.7% correct judgments), McNemar's $X^2(1) = 16.413$, $p < 0.0001$. In response to a question about whether it was easier to classify first and second themes from the scores or the recorded excerpts, 61% (27/44) of the participants indicated that the recording-based task was easier than the score-based task.

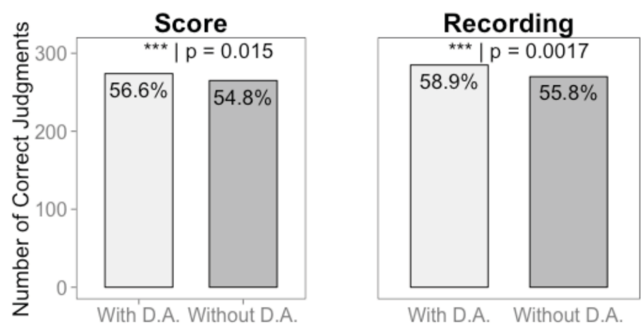


Figure 2. Participants perform slightly better in situations when the music contains articulation and dynamic information ("With D.A.") than when the music does not contain this information ("Without D.A.") for both the score-based and recording-based judgments.

C. Criteria Used in Decisions

After the participants finished the two tasks, they were asked to rank order the criteria they used to make their determinations. In addition to the criteria from the corpus study, five blanks were provided so participants could list their own criteria. The results are shown in Table 1. Of the five features expected to differ between first and second themes, rhythmic smoothness (an indication of nPVI) was considered to be the most important feature to aid in discrimination between first and second themes. Articulation and dynamics were consistently rated on the lower half of the list of criteria. This is not surprising, as only half of the samples seen by participants contained dynamic and articulative information (the "With DA" condition, as opposed to the "Without DA" condition). The category marked "Other" was also considered to be important in making the determinations of first and second themes. Participants listed features such as the note density (texture), harmonic complexity, melodic range, cadence type, phrase length, type of meter (simple or compound), speed, and difficulty (virtuosity) as features that helped make the determinations between first and second themes. These additional features are listed in Table 2.

Table 1. Rank-ordered criteria used to determine which theme from a pair comes first or second in the music. The mean importance of each criterion is shown, where the smallest number means it is considered to be the most important (closest to rank number 1).

Criteria	Scores	Recordings
Articulation	3.56	3.80
Dynamics	3.47	3.13
Durational Pace	2.68	2.95
Interval Size	4.15	4.12
Rhythmic Smoothness	2.22	2.52
Other	2.77	2.23

Table 2. Additional criteria listed by participants as features that helped them perform the discrimination task, with the number of mentions for each feature. Criteria with no mentions in each task are italicized.

Scores		Recordings	
Variations	1	Variations	1
Texture	4	Texture	2
Sequences	2	Sequences	2
Elaboration	1	Elaboration	1
Declamatory Style	1	Declamatory Style	1
Harmonic Complexity	4	Harmonic Complexity	5
Difficulty	2	Difficulty	1
Other musical markings (e.g. pedal)	2	*Other musical markings (e.g. pedal)*	*0*
Cadences	2	Cadences	1
Modulation	2	Modulation	2
Harmonic Rhythm	2	Harmonic Rhythm	1
Melodic Range	1	*Melodic Range*	*0*
Chromaticism	3	Chromaticism	1
Mode	1	Mode	2
Melodic Smoothness	1	Melodic Smoothness	1
Tempo/Speed	*0*	Tempo/Speed	3
Meter	*0*	Meter	1
Phrase Length	*0*	Phrase Length	1
Phrase Completeness	*0*	Phrase Completeness	1

D. Listener Attributes

A few *post-hoc* tests were conducted regarding characteristics of the listeners. The first consideration regarded the amount of music theoretical training in which the participants had engaged. The mean years of theory training in the participant group was 3.66 years. It was thought that music theory graduate students and upper level undergraduate music theory majors might perform better on the tasks. The participants were accordingly split into two groups. The “Less Theory” group (n = 34) consisted of those who had less than 4 years of music theory training, whereas the “More Theory” group (n = 10) consisted of those who had 4 or more years of training. As expected, the “More Theory” group performed better on the tasks (59.5% correct) than did the “Less Theory” group (55.6% correct), McNemar's $X^2(1) = 173.65$, Bonferroni-corrected $p < 0.001$. The participants with higher musical training performed better on the recording task (61.8% correct) than on the score-based task (57.3% correct), Bonferroni-corrected $p < 0.05$. The participants with less training also performed better on the audio task (56.0% correct) than on the notation task (55.2% correct), although the effect size is small, Bonferroni-corrected $p < 0.05$.

The second *post-hoc* test examined whether or not participants with training in sonata theory (n = 23) would do better on the task than participants without training in sonata theory (n = 21). Although skewed in the expected direction, the difference was not statistically significant, Bonferroni-corrected $p = 0.126$.

It might be expected that listening to and engaging with music on a regular basis probably also contributes to perceptual judgments about first and second themes. Accordingly, two further tests were carried out. Participants were categorized by the amount of instrumental training they had received. The average amount of training in the participant groups was 9.4 years of instrumental instruction. The participants were divided according to a mean split into a “High Training” (10 or more years of training) and a “Low Training” category (n = 24 and n = 20, respectively). The groups performed equally well on the tasks, Bonferroni-corrected $p = 2.83$).

Finally, participants were split into two categories based on how much time they spent listening to classical music. A mean split was performed (mean = 4.33 hours), resulting in “High Listening” (n = 31) and “Low Listening” (n = 13) categories. Those in the high listening category performed significantly better on the task (59.4% correct) than did those in the low listening category (55.3% correct), McNemar's $X^2(1) = 76.169$, Bonferroni-corrected $p < 0.00001$.

Of course, we must be cautious in interpreting these results. There is a certain arbitrariness in dividing a continuous variable into categories, even when well-intentioned. Moreover, many of the groups performed similarly to each other, despite the statistical significance between the groups. Two important trends do emerge, however. Those with more theory training and those who regularly listen to classical music performed the best on the discrimination tasks. This follows basic intuition about musical training.

VI. IMPORTANCE OF HARMONY

A modified version of the perceptual task was conducted after the data analysis from the original perceptual study was carried out. Recall that when participants were asked to describe the criteria they used to make their decisions, many participants indicated that they used harmonic considerations. In this modified task, the same sample of musical works was utilized. However, the harmony was removed so that the musical excerpts consisted of only a single melodic line (see Figure 3). The task protocol remained the same.

Twelve participants completed this modified task, none of whom had taken the original perceptual task. The average age was 20 years (range from 18-25), and 58% were female (7/12 participants). The participants had an average of 4.6 years of music theory training (range from 1.5-14 years) and an average of 8.5 years of instrumental or vocal training (range from 1-14). The results are not consistent with the hypothesis that participants can differentiate first and second themes, with 48% accuracy in judgments, p = 0.408. There was no difference in performance for music with or without dynamics and articulation, $X^2(1)$ = 0.329, p = 0.566. Participants performed equally well on the score-based and recording-based tasks, $X^2(1)$ = 0.291, p = 0.589. A median split was performed to divide participants into groups with more and less music theory training. Both of these groups performed equally well on the tasks, $X^2(1) = 0.307$, $p = 0.580$. The results seem to indicate that participants are not able to distinguish first and second musical themes without access to the underlying harmony.

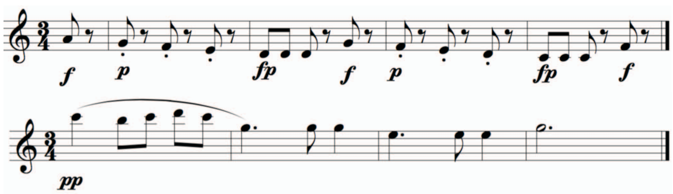



Figure 3. First theme (top line) and second theme (bottom line) from Dvořák's *Waltzes, Op. 54, No. 3.*

VII. CONCLUSION

The goal of the current study was to determine whether people could differentiate between a pair of musical themes using visual and auditory factors. Overall, participants were able to use structural musical features to differentiate first and second themes from the same musical work; they could do this both from sound recordings and from musical scores. In general, participants performed better in the conditions where dynamics and articulation were available than in conditions without these features; however, the performance across these two conditions was similar. In general, performance was similar across many of the conditions of the study. For example, in both the score-based task and recording-based task, accuracy in judgments was consistently around 57-60%. This is significantly different from chance, but is small in magnitude, indicating that the task was difficult. Participants with more theory training and those who regularly listen to classical music performed better on the discrimination tasks.

Score-reading and close-listening are related, but distinct musical skills. Despite the nearly identical performance on the score-based and recording-based tasks, participants seem to indicate that the recording-based task was easier than was the score-based task (61% indicated that the recording-based task was easier, but only 18% indicated that the score-based task was easier).

The corpus study carried out by Warrenburg and Huron (2015) established that there are indeed broad structural differences between first and second themes in classical Western instrumental repertoire—consistent with observations made by music scholars. The question addressed in the current study is whether musician listeners are sensitive to these differences.

On the one hand, the enhanced performance of participants who regularly listen to classical music suggests that an ability to discriminate first and second themes might arise through simple exposure involving implicit learning. However, the enhanced performance of participants with theory training suggests that the ability to carry out the task may depend primarily on explicit theoretical knowledge. Those participants who listen regularly to classical music are also likely to have received greater theoretical training. In any event, the small effect size indicates that the task of discriminating first and second themes is far from easy. Sensitive listeners may indeed be aware that a given theme is likely to appear first. However, the current study suggests that it is only a small component of music perception, even for musician listeners.

REFERENCES

Barlow, H., & Morgenstern, S. (1948). *A dictionary of musical themes*. New York, NY: Crown Publishers.

Burnham, S. (2002). Form. In T. S. Christensen (Ed.), *The Cambridge history of Western music theory* (pp. 880-906). Cambridge: Cambridge University Press.

Caplin, W. E. (1998). *Classical form: A theory of formal functions for the instrumental music of Haydn, Mozart and Beethoven.* New York: Oxford University Press.

Caplin, W. E., Hepokoski, J., & Webster, J. (2010). *Musical form, forms, & Formenlehre: Three methodological reflections* (2nd ed.). Pieter Bergé (Ed.). Leuven, Belgium: Leuven University Press.

Churgin, B. (1968). Francesco Galeazzi's description (1796) of sonata form. *Journal of the American Musicological Society*, *21*(2), 181-199.

Cook, N. (2002). Epistemologies of music theory. In T. S. Christensen (Ed.), *The Cambridge history of Western music theory* (pp. 78-105). Cambridge: Cambridge University Press.

Dunsby, J. (2002). Thematic and motivic analysis. In T. S. Christensen (Ed.), *The Cambridge history of Western music theory* (pp. 907-926). Cambridge: Cambridge University Press.

Hepokoski, J. A., & Darcy, W. (2006). *Elements of sonata theory: Norms, types, and deformations in the late-eighteenth-century sonata*. New York: Oxford University Press.

Huron, D. (1993). *The Humdrum Toolkit: Software for Music Research*. Center for Computer Assisted Research in the Humanities.

Koch, Heinrich Christoph. (1983). *Introductory essay on composition: The mechanical rules of melody, sections 3 and 4.* (N. K. Baker, Trans.). New Haven: Yale University Press. (Original work published 1793).

Kollmann, A. F. C. (1973). *An essay on practical musical composition: According to the nature of that science and the principles of the greatest musical authors*. New York: Da Capo Press. (Original work published 1799).

Low, E. L., Grabe, E., & Nolan, F. (2000). Quantitative characterizations of speech rhythm: syllable-timing in Singapore English. Language & Speech, 43 (4), 377–401.

Marx, A. B. (1845). *Die Lehre von der musikalischen Komposition, praktisch-theoretisch*. Vol. 3. Leipzig: Breitkopf und Härtel.

Newman, W. S. (1963). *The sonata in the classic era*. Chapel Hill: University of North Carolina Press.

Rosen, C. (1971). *The Classical style: Haydn, Mozart, Beethoven.* New York: Viking Press.

Schellenberg, E. G., Corrigall, K. A., Ladinig, O., & Huron, D. (2012). Changing the tune: listeners like music that expresses a contrasting emotion. *Frontiers in Psychology*, *3*(574), 1-9.

Warrenburg, L. A., & Huron, D. (2015, August). *On contrasting expressive content between primary and secondary musical themes.* Poster presented at the Society for Music Perception and Cognition, Nashville, TN.

Using eye tracking to investigate intense music listening

Elke B. Lange[1], Fabian Zweck[2], and Petra Sinn[3]
[1]Max-Planck-Institute for Empirical Aesthetics, Frankfurt, Germany
[2]Bavarian Broadcasting, Munich, Germany
[3]University of Potsdam, Potsdam, Germany

Experiencing music can be accompanied by the state of being absorbed by the music, a state in which individuals allow the music to draw them into an intense listening experience. We investigated whether being absorbed by music might be related to eye movement control, pupil size and blinking activity. All three measurements index cognitive workload. If they relate to felt absorption, they might pick up cognitive processing as part of state absorption. We presented musical excerpts to listeners in two Experiments. In Exp. 1 we instructed participants to fixate centrally, measuring microsaccades, miniature movements of the eyes during fixational movements. In the free viewing task of Exp. 2 we measured larger, eye movements, called saccades. Eye movements, pupil dilation and blink rate were recorded. Musical excerpts were selected from a broad range of styles. The listeners rated each excerpt for felt absorption, valence, arousal, preference and familiarity. Acoustic properties of the music were analyzed using the MIR toolbox. We used linear mixed models to predict the subjective ratings by the physical properties of the eye. In Exp. 1 we found clear evidence for microsaccade and blink rate to predict absorption and liking. Predicting valence included an additional effect of pupil size. Knowing and arousal were not reliably related to our measurements. In Exp. 2 fixational microsaccades were not expected to be effective as eyes were moved freely. Here, some of the effects shifted to mean saccadic amplitude. Acoustic properties contributed to predictions of subjective ratings as well in both Experiments. Our results are important for two reasons. First, objective measurements can predict subjective states during music listening. Second, the construct of state absorption might be related to cognitive processing. To sum, eye tracking seems to be a fruitful tool to investigate musical processing during listening to music.

Extracting the Musical Schemas of Traditional Japanese Folk Songs from Kyushu District

Akihiro Kawase,[*1]

[*]*Faculty of Culture and Information Science., Doshisha University, Japan*
[1]akawase@mail.doshisha.ac.jp

ABSTRACT

A. Background

We have sampled and digitized the five largest song genres within the music corpora of the "*Nihon Min'yo Taikan*" (1944-1993, NHK) consisting of 1,794 Japanese folk song pieces from 45 Japanese prefectures, and have clarified that the differences in the musical characteristics almost match the East-West division in the folkloristics from a broader perspective (Kawase & Tokosumi, 2011). However, to conduct more detailed analysis in order to empirically clarify the structures by which music has spread and changed in traditional settlements, it is necessary to expand the data and do comparisons based on the old Japanese provinces (Kawase, 2015).

B. Aim of the study

The main purpose of this study is to extract the pitch transition patterns from pieces of traditional Japanese folk songs, and by making comparisons with Koizumi's tetrachord theory (Koizumi, 1958) in order to make the regional classification by the tendency in pitch information.

Influenced by the methods of Western comparative musicology, the Japanese musicologist Fumio Koizumi conceived of a scale based not on the octave unit but rather on the interval of a perfect fourth, and has developed his tetrachord theory to account traditional Japanese music (Koizumi, 1958). The tetrachord is a unit consisting of two stable outlining tones called *kaku-on* (nuclear tones), and one unstable intermediate tone located between the nuclear tone. Depending on the position of the intermediate tone, four different tetrachords can be formed (see Table 1 and Figure 2).

Table 1. Koizumi's four basic tetrachords.

Type	Name	Pitch intervals
I	*Min'yo*	minor third + major second
II	*Miyako-bushi*	minor second + major third
III	*Ritsu*	major second + minor third
IV	*Ryu'kyu*	major third + minor second

C. Methods

As a case study, we extracted the musical notes for works included in the "*Nihon Min'yo Taikan*" (Anthology of Japanese Folk Songs) (1944-1993, NHK), which is a collection of catalogued scores for Japanese folk songs. We digitized the entire scores from each province in the Kyushu district (geographically located in the southern part of Japan) into MusicXML file format. In the experiment, (a) 89 Buzen song pieces, (b) 99 Bungo song pieces, (c) 46 Chikuzen song pieces, (d) 29 Chikugo song pieces, (e) 163 Hizen song pieces, (f) 150 Higo song pieces, (g) 107 Hyuga song pieces, (h) 77 Satsuma song pieces, (i) 116 Ohsumi song pieces, (j) 11 Tsushima song pieces, (k) 9 Iki song pieces were selected. In total there were 382,653 tones in the sample of 896 songs for the 11 provinces (e.g. Figure 1).

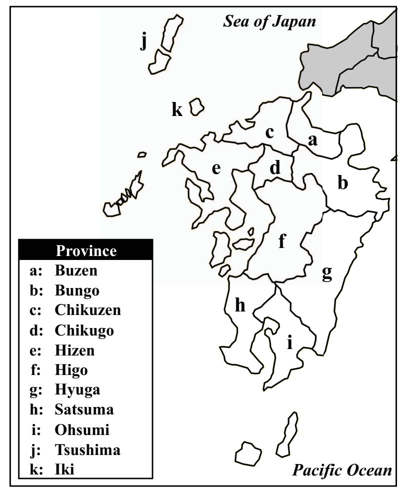


Figure 1. Geographical divisions of the Kyushu district under the old province system.

In order to digitize the song pieces, we generated sequences of notes as follow. In the classification experiment of folk music using hidden Markov models, Chai and Vercoe represent melodies in several ways. They suggested that the "interval representation" is more compatible with human perception, as people employ interval information when they memorize, distinguish or sing a melody (Chai & Vercoe, 2001). Here we adopt their interval representation by digitizing each note in terms of its relative pitch. By subtracting the next number (pitch), it is possible to generate a pitch sequence $T = (t(1), t(2), t(3), \dots , t(i), \dots , t(n))$ that carries information about the pitch interval to the next note, where, $t(i)$ are the notes i, and n is the number of elements in a sequence T.

We quantitatively analyzed its pitch transition patterns using N-gram modeling, and conducted a hierarchical cluster analysis based on the frequency of occurrences of the tetrachords to see the differences in each province.

D. Results

Results (will be reported at the conference in more detail) indicate that pitch transition patterns of Japanese folk songs of Kyushu district have an overwhelming tendency to either form perfect fourth pitches or return to the initial pitch within two or three steps. In addition, a difference among provinces is observed in comparing frequencies of generating the tertrachords identified by F. Koizumi.

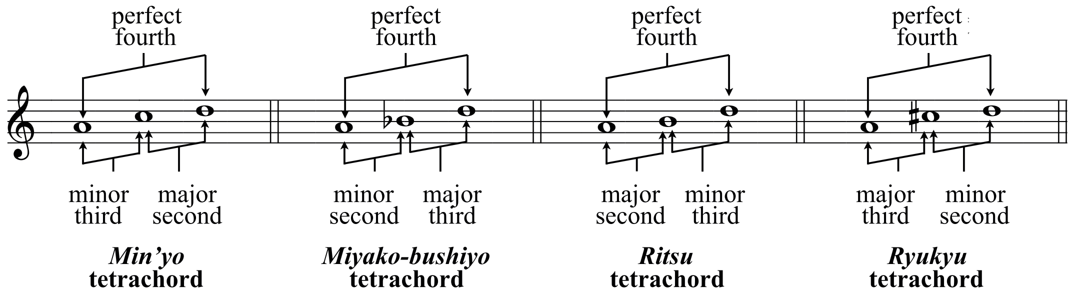


Figure 2. Example of four different types of F. Koizumi's tetrachord when taking A-D as *kaku-on* (nuclear tones).

The dendrogram shown in Figure 3 is the result of the hierarchical cluster analysis, and the result of applying the Ward method to the extracted probabilities for each province. As a result, we constructed the possibility that the melodic features in the Kyushu district spread by land and sea routes based on quantitative analysis (see Figure 4).

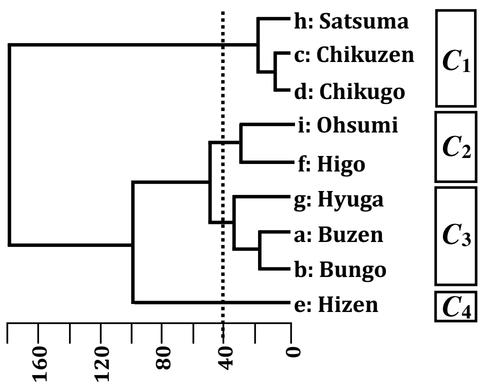


Figure 3. Dendrogram based on transition probabilities of tetrachords for each province.

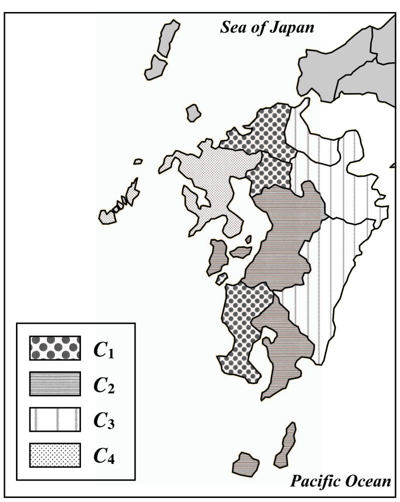


Figure 4. Classified results plotted on the map.

E. Conclusions

This study focused on melodies of endangered Japanese folk songs, and examined the characteristics of its schema from a quantitative perspective. It is possible to clarify the structural commonalities and differences between areas by conducting analysis on musical corpora nationwide including folk pieces from neighboring regions, which has not really been pursued in existing humanities research fields. We believe this will empirically clarify the musical culture phenomena by which music spreads and changes.

ACKNOWLEDGMENT

This work was mainly supported by the Japanese Society for the Promotion of Science (JSPS) Grants-in-Aid for Scientific Research (15K21601) and the Suntory Foundation Research Grants for Young Scholars.

REFERENCES

Chai, W., & Vercoe, B. (2001). Folk music classification using hidden Markov models. In *Proceedings of International Conference on Artificial Intelligence 2001*.

Kawase, A., & Tokosumi, A. (2011). Regional classification of traditional Japanese folk songs. International Journal of Affective Engineering, 10(1), 19-27.

Kawase, A. (2015). Quantitative analysis of traditional Japanese folk songs in Kyushu region. In *Proceedings of the Fifth Annual Conference of Japanese Association for Digital Humanities 2015.* Japan: Kyoto University.

Koizumi, F. (1958). *Nihon Dento Ongaku No Kenkyu 1* (Studies on traditional Music of Japan 1). Tokyo: Ongaku no Tomo Sha.

Nihon Hoso Kyokai. (1944-1993). *Nihon Min'yo Taikan* (Anthology of Japanese Folk Songs). Tokyo: NHK Publishing.

An Investigation on the Perception of Regional Style of Chinese Folk Song

Cong Jiang[*1]

[*]College of Music, Capital Normal University, China
[1]conger108@yahoo.com

ABSTRACT

Folk songs in the same tune families are quite popular in Chinese music. Musicologists in Chinese traditional music studied these folk songs from historical and music analysis perspectives, but few studied on listeners' perception of folk songs. This study used one "*tongzong*"(similiar to tune family) - Molihua(Jasmine) - including 4 versions from different provinces, to analyze the similarity in motifs, melody contour, rhythm and lyrics, and also to investigate the perception of regional styles by music and non-music students. It was found that perception of regional style depended on both popularity of folk songs and individual cultural background of students; music students responded more accurately than non-music students. However, only few students were very confident with their responses. It was discussed that fluctuation of melodic contour and tempo may contribute to the perception of regional style in "*tongzong*" families.

I. INTRODUCTION

In Chinese folk songs, there are series of folk songs which can be categorized into different tune families -- which in Chinese called "*tong zong*" that means the songs in a tune family have the same root or origin (Feng, 1998). According to the musicologists in Chinese traditional music, these folk songs are generally categorized according to their melodies, lyrics, modes or tonal structures, or even names of the tunes (in Chinese *qu pai*). So for the concept of "*tongzong*" and tune families, there are some differences: the concept of "*tongzong*" is more general which includes the similarities of lyrics and names of the tunes besides musical features; while the concept of tune family is relatively narrow which only involves musical similarities, the same as some western researchers mentioned (Volk & van Kranenburg, 2012).

A "*tongzong*" category usually involves songs with similarities in both tunes and lyrics, or songs with similarities only in tunes or only in lyrics, or songs with similarities in structures (Feng, 1998). Although there are some slightly different opinions held by scholars - who argued that only songs with tune similarities should be involved in a "*tong zong*"(Xu, 2001), this study will not discuss the rationality of "*tongzong*" categorization in Chinese folk songs, and just take a broader concept of "*tongzong*" . In a "*tongzong*" category, although folk songs have similiarities in music or lyrics, they still have another feature -- regional music style. Because folk songs were brought to different provinces through immigration or wars, etc., they were characterized with local musical features. Then there might be a conflict in grouping "*tongzong*" and regional style. If folk songs are categorized into different "*tongzong*" families, can listeners tell the regional style of folk songs? If they can, what are the clues for listeners? Are these clues separated from those similar features or are they mixed?

So this study selected one "*tongzong*" category -- "*Molihua*" (Jasmine), one of which melody was used in Giacomo Puccini's opera *Turandot*, is now found at least 11 versions. The similarity of melody, rhythm and lyrics were analyzed, and presented to listeners to investigated whether they could tell regional style of folk songs or not.

II. *MOLIHUA IN MUSICOLOGY*

The folk song *Molihua* (Jasmine) is a well-known ditty in China. The earliest version called *Xianhua diao* (flower tune). Although the origin of the song is still unknown, the earliest record of lyrics was found in a traditional Chinese drama script *Zhuibaiqiu*, in which there were two pieces of lyrics in two ditties called *Huaguqu* that was similar to the lyrics of *Xianhua diao,* around Emperor Qianlong and Jiaqing period in Qing dynasty. The earliest record with Gongche Pu ("Chinese pitched notation in which seven Chinese characters are used to indicate seven pitches", Zhang, 2009) was found in *Xiaohuiji* in Emperor Daoguang period in Qing dynasty (1821 or 1837). The most popular version nowadays is from Jiangsu province, which is also well-known in the world through Giacomo Puccini's opera *Turandot.*

A. Historical changes

The "original" version *Xianhua diao* consists of 4 phrases, the mode is zhi mode with five tones in the scale, the tempo is not too fast, and the second verse of lyrics is quite similar to the Jiangsu version nowadays (Xiong, 2013). It is said that this version was also transcribed by Hefang in Liuhe, Jiangsu province in 1942. And other scholars also found quite the same result with *Huaguqu* in the Chinese drama script *Zhuibaiqiu*, and *Huanguqu* was a kind of folk music from Fengyang, Anhui province (Yang, 2014). Thus, it is difficult to say whether *Molihua* was originally from Jiangsu or Anhui province, and the song might not be only a ditty, but a part of secular opera or drama at the very beginning.

For its widely spreading in China, it might due to folk artists' wandering in different places for living - they went to the north, the west and the south to give performance. *Zhuibaiqiu* could be one of their repertoire, and *Huanguqu* played an important role in the drama, so people listened it for several times and sang it for generations. It is now found about 11 versions in Jiangsu, Henan, Hebei, Heilongjiang, Liaoning, Ningxia, Shanxi, Hubei, Gansu, Guangxi, Shaanxi Provinces (Feng, 1997). In a new place, the original music and songs were like seeds absorbing nutrients from local culture, and bloomed new versions with regional style. In some versions, there are the same lyrics starting with "hao yi duo molihua" ("how delightful this branch of fresh flowers, translated by John Barrow, 1804, Gong, 2013), for instance, the versions of Jiangsu, Hebei, Shanxi, Heilongjiang, and Liaoning (some modal particles were added); some other versions changed the lyrics to narrate other love stories, such as Ningxia version.

B. Analysis of four versions

This study only involved four versions from Jiangsu, Hebei, Heilongjiang, and the northeast area of China, which audio recordings were available. Only Heilongjiang version was sung in male voice, the other three versions were sung by women. The common features of the four versions are:

- there are 4 phrases;
- the rhythm is 2/4 beat (except Heilongjiang version: 4/4 beat);
- the scale is pentatonic (except Hebei version: hexatonic);
- in zhi mode (ended on *sol* in movable-do system).

The differences in melodic contour and register, motif, and tempo of the four versions are compared here.

1) Melodic contour & register

The melodic contour of each song recording was analyzed by Praat (version 5.4.06), which generated a pitch list of each phrases in the song, and then the contour was drawn in Excel with the pitch list. So the melodic contour in this study is not a perceived one, but an algorithmic contour.

The melodic contour presented here only involve the first phrase "hao yi duo molihua", which is repeated once in each version. Jiangsu and Heilongjiang versions (see figure 1a and 1b) are more melodic, smooth and legato; while Hebei and Northeast versions (see figure 1c and 1d) are with more intervals, because the lyrics in these two versions contain more modal particles.

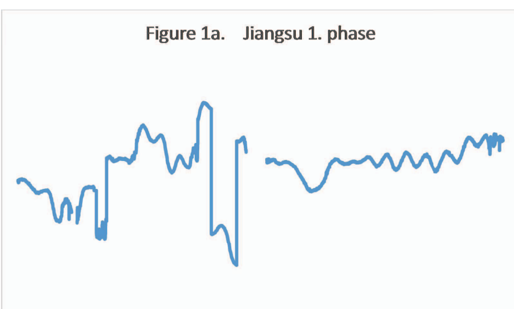


Figure 1a. Melody contour of first phrase in Jiangsu version

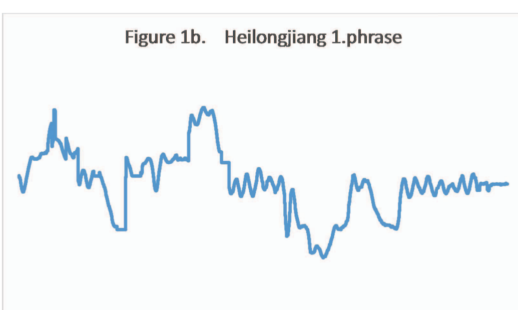


Figure 1b. Melody contour of first phrase in Heilongjiang version

For Jiangsu and Heilongjiang versions, the two melodic contours are generally quite similar, and they are only different at the beginning and later part. Jiangsu version covers an interval of minor seventh. It starts more flat, and then gradually goes up and then comes down, so the contour is like an arch-shape; while Heilongjiang version covers an interval of minor sixth. It starts more dramatically and goes up first, and then comes down a little, and rises again and decreases, so the contour is like two-humped camel.

For Hebei and Northeast versions, both of them cover an interval of minor sixth, and have some modal particles in lyrics. Hebei version fluctuates also twice, but the second peak of contour comes a little bit later; while the generated melodic contour of Northeast version is not so clear, because the singing is with more staccato, and the contour is similar to Heilongjiang version, but flatter with only one peak in the contour.

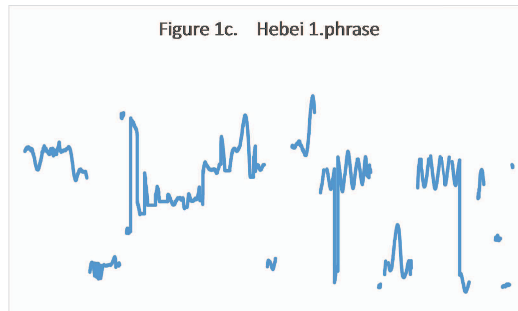


Figure 1c. Melody contour of first phrase in Hebei version

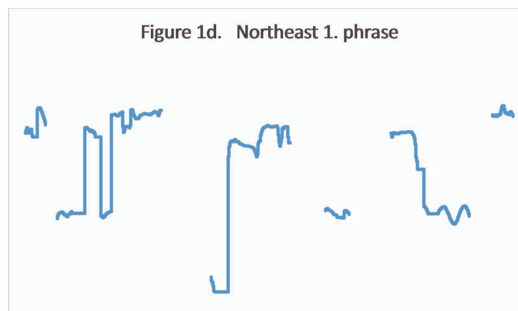


Figure 1d. Melody contour of first phrase in Northeast version

Using Kendall nonparametric test to calculate correlations of four melodic contours, it showed that there are significant correlation between Jiangsu and Heilongjiang version ($r=0.078$, $p<0.05$), Jiangsu and Hebei version ($r=0.124$, $p<0.01$), Jiangsu and Northeast version ($r=0.116$, $p<0.01$), Hebei and Heilongjiang version ($r=0.203$, $p<0.01$), and Heilongjiang and Northeast version ($r=0.101$, $p<0.01$).

Motif

Using interval vector of the music set theory to describe the common feature of the contour, it is found that the common motif of four versions is in [03583] intervallic relationship: a minor third goes upward, followed by an upward major second, and then an upward minor third, and ends with a downward perfect fourth. It sounds like "mi - sol - la - do - sol" in movable-do system.

This intervallic relationship could be presented directly in the melody, or indirectly in the melodic structure, or used simplified in the melody. For instance, Heilongjiang and Hebei version directly show the motif in the melody in the first phrase; Jiangsu version shows the motif in its melodic structure, that is, there are some tones added in the motif "frame" to decorate the melody; and the motif in Northeast version is shorted to [0250] (sol - la - do - sol) without minor third in its melodic structure.

Tempo

The duration of four audio recordings are quite different (see table 1). Jiangsu version is the fastest, while Heilongjiang and Northeast versions are the slowest. There is no modal particles in the lyric of Jiangsu version, so six words correspond 12 notes; two words were added at the end of the lyric of Heilongjiang version, totally 8 words correspond 16 notes; three words were added respectively in the middle and at the end of the lyric of Hebei version, totally 9 words correspond 18 notes; and two words were added respectively in the middle and at the end of the lyric of Northeast version, totally 8 words correspond 11 notes, and besides, another 5 model particles prolong or complement the phrase for another 2 beats.

Table 1. Tempo of four versions

	Jiangsu	Heilongjiang	Hebei	Northeast
Duration	3.25s	6.88s	7.14s	4.45s
No. of tones	12	16	18	11
No. of beats	4 (2/4)	6 (4/4)	8 (2/4)	4 (2/4)
Tempo (bpm, ca.)	73-74	52-53	67-68	53-54

Thus, considering melodic contour, motif and tempo, Jiangsu version may sounds delicate because of its fast tempo and decoration tones in the motif; Heilongjiang version may sounds unrestrained because of its dramatic contour and slower tempo; Hebei version may sounds witty because of its modal particles and more tones in the melody; and Northeast version may sounds melodious because its slower tempo and less changed melodic contour.

III. AIM

From music analysis, it is easy to find the similarities of four version *Molihua*s in melodic contour, motif, scale, mode, rhythm, tempo and lyrics. However, it is interesting to see that "*tongzong*" songs also have regional styles, and these characters are also mixed with different musical elements. If people can tell a set of "*tongzong*" songs based on lyrics, motif, or melodic contour, etc., can they also distinguish the regional styles of the songs? What aspects influence their judgment? And what relationship between the features of "*tongzong*" category and the features of regional style? This study tried to investigate these questions in perception of listeners.

IV. Method

A. Materials

The four versions were collected from published recordings. Although many singers sang and recorded the songs, but not all of them sounds very "regional" or "local". So it would be better to find some versions that were sung by native singers of certain region. Here is the list of recording and singers:

- Jiangsu version, Xiufang Ju (1934-), from Jiangsu province;
- Heilongjiang version, Song Guo (1931-), from Shenyang province, studied and lived in Heilongjiang province for many years.
- Hebei version, Lan Si, from Anhui province and a folk vocal singer.
- Northeast version, the singer is unknown.

B. Participants

Participants involved both music students and non-music students. There were 117 music students, major in instrument, vocal, music education and recording; and 81 non-music students, major in mathematics, physics, biology, chemistry, computer science, architecture. All of them were freshmen in the university, in their 18 or 19 years old, and they are from more than 20 provinces in different parts of China.

C. Task

Each version were presented only the first verse of the song, and played twice for each. Students were asked to answer where the song originally comes from, and their judgment could be made according to their musical experience or based on their intuition. The answer of the origin should be as precise as possible, for instance, it would be better to state "Jiangsu province" than "east part of China".

Besides, they should also answer on which they depended to make the judgment, either on lyrics, melody (for both music and non-music students) or even instrumentation (only for music students); and evaluate how confident they were about their answers on a 5-point Likert-scale.

V. Result

All the perceived region were transmitted into numbers which indicate the distance between the original region and perceived region, and the distance was not in kilometers, but in the number of provinces. For instance, if a student answered Jiangsu province to Jiangsu version, then he would get 1 point; if a student answered Hunan province to Jiangsu version, then it was checked on the map of China that people should go across Jiangsu, Anhui, Hubei or go across Jiangsu, Anhui, Jiangxi to get Hunan, so the student will get 3 points.

A. Distribution of perceived origins

Generally speaking, music students responded more correctly than non-music students, and the incorrect answers of music students did not scatter as widely as those of non-music students (see figure 2a & 2b).

For Jiangsu version, which is the most popular version in China, there were 114 valid answers of music students, and 57.89% answered correctly; all the answers of 81 non-music students were valid, only 37.84% answered correctly, and the percentages of incorrect perceived regions in other four different distances were more than music students.

For Hebei version, there were 114 valid answers of music students, and 40.35% answered correctly, which was much more than the correct percentage 10.53% of 80 valid answers of non-music students. However, it was interesting to see that both music and non-music students (44.74%, 64.47%) perceived the regional style which was a little bit far away from the original region.

The same situation with Heilongjiang version, and the percentage of perceived regional style in further region was even increased in both music and non-music students (66.38%, 79.49%).

For Northeast version, the perceived regional style was even further from the original region, and incorrect answer also got more.

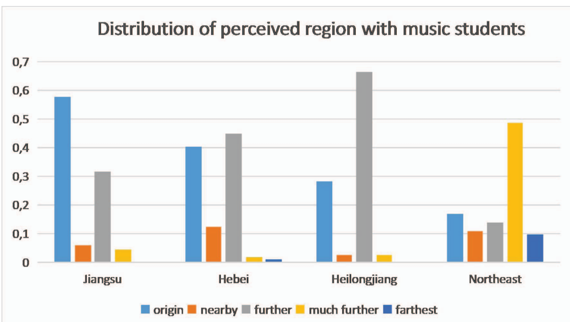


Figure 2a. Distribution of perceived region with music students

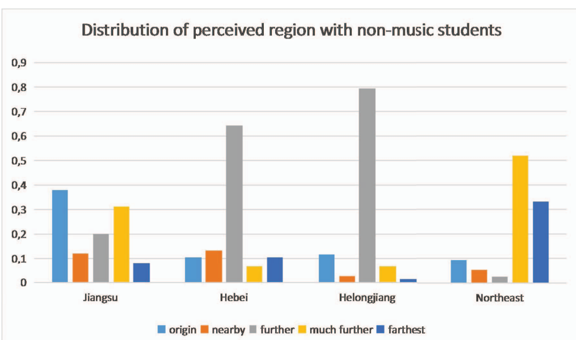


Figure 2b. Distribution of perceived region with non-music students

B. Distance between original region of songs and hometown of students

The distance between original region of songs and hometown of students was also calculated with number of provinces between them. Students may be familiar with music culture of their hometown, but feel strange with music culture in other places. Figure 3a and 3b show the accuracy of students who are from different distance to the original region of four versions. For instance, more than 60% of music students from Hebei province responded correctly, and over 40% of music students from the provinces neighboring Hebei province responded correctly, etc.

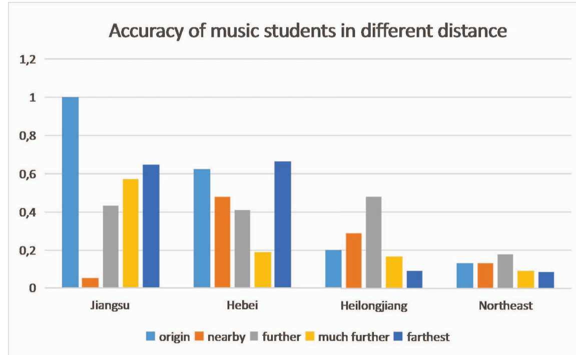


Figure 3a. Accuracy of music students from different distance

From figure 3, the percentages of music and non-music students from Hebei, Heilongjiang or Northeast are quite the same with the percentages of correct answers to these three versions, so regional culture background may play an important role in judgment making. While, the responses to Jiangsu version were independent from regional culture background that indicates its popularity.

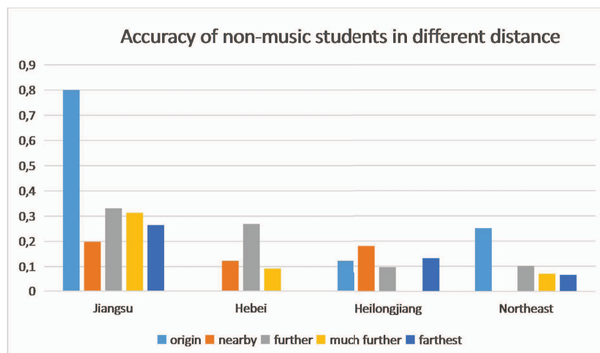


Figure 3b. Accuracy of non-music students from different distance

C. Reference and confidence of responses

Most music students took melodies as reference, although instrumentation was also offered as another option for them and only few students took it as clue for judgment. While, non-music students were only asked to report whether they made judgments on melody or lyrics. Over 70% students in both groups reported that melodies were the clues for the judgments. As for the confidence, only less than 10% students were very confident with their response to any versions. It was interesting to see that music students did not show higher confidence than non-music students.

Considering both region judgments and rate of confidence, students from different distance to the original regions of folk songs were compared: music students from different distant regions showed significant differences in their responses to the four versions ($p<0.01$). Using SNK post hoc multiple comparison to find homogeneous subsets, for Jiangsu version, there was only one homogeneous subset; but for Hebei version, the students from furthest regions were separated from the other four groups; for Heilongjiang and Northeast versions, the students from much further and furthest regions were separated from the other three groups.

Non-music students also presented the significant differences among different distant regions (p<0.01). For Hebei and Heilongjiang version, the students from furthest regions were separated from the other four groups; for Northeast versions, the students from further, much further and furthest regions were separated from the other two groups, and the latter two groups were in the same homogeneous subset.

VI. Conclusion

This is a pilot study to explore the perception of regional style of "*tongzong*" folk songs. This study took a broad sense of "*tongzong*" category which involves similarities in both music and lyrics, and folk songs were also chosen with singing version, so it was assumed that listeners could categorize these songs according to features either in music or in lyrics. That is, it was taken as a default that when students heard the first phrases of four "*Molihua*" versions, they could tell the four versions shared the same topic. Based on this assumption, it was investigated whether students could tell the region style of each version.

From the accuracy of students' responses on region styles, it seems that familiarity of songs plays an important role, because Jiangsu version is so popular for nearly every Chinese. And influence of such familiarity may be even stronger than the familiarity of culture in their hometown which could be more implicit as part of home identity. Just as the students from the province where is the origin of the folk song, they got higher percentage of accuracy than the students from other areas.

The similar low confidence of both music students and non-music students may be not in the same situation. Although the music students rated low on their confidence, but it seems that they have smaller "hesitant area" than non-musical students, that is, they were quite sure or clear about whether they know the region styles of songs or not. While some of non-music students tried to seek the solution with modal particles in the lyrics which also present some local features.

How music elements play roles in their judgments? It is a little bit difficult to answer this question in this study. In the musical analysis, the most common feature of the four "*Molihua*" versions is the motif, which is presented directly in the melodies or presented indirectly in melodic structures; and also rhythm, which is in the same or in a comparable meter. As former studies mentioned, characteristic motif is the most important for tune family categorization (Volk & van Kranenburg, 2012; Cowdery, 1984), while lyrics could be hardly the main reason for categorization (Volk & van Kranenburg, 2012). Motif could also play an important role in "*tongzong*" family categorization either as local or global features, but this study did not deal with this issue. However, this can be investigated in further studies with non-lyric versions.

Different from the study by Volk and Kranenburg, in which they tried to find out the musical similarities of 360 folk song melodies in 26 tune families, this study focused on whether listeners could distinguish regional style of "*tongzong*" family. That is, characteristics of "*tongzong*" family and those of regional style are different categorizations. In music analysis, similarities of motif, melody contour and comparable rhythm are usually mentioned in the tune families or "*tongzong*" families (Selfrige-Field, 2007; Volk & van Kranenburg, 2012; Cowdery, 1984; Feng, 1998; Xu,2001); while, styles in different regions or in different nations often refer to fluctuation of melody contour, modes and rhythm-tempo in Chinese folk song analysis (Yang, 2002). Thus, perception of region style in "*tongzong*" family is a crossover study. It would be better to compare with the perception of folk songs in certain region, so that it would be clear that whether the common regional features of some version in the "*tongzong*" category work in the same way or not. However, such research is relative rare at the moment.

In consideration of musical analysis of four versions, despite of common features in motif, rhythm and modes, fluctuation of music contour and tempo were giving direction of region judgment. Furthermore, the strong connection between dialects and tunes of folk songs can also play an important role. These have to be investigated in the future studies.

ACKNOWLEDGMENT

This study is a part of project "Implicit musical knowledge of Chinese students", sponsored by the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry, 2014

REFERENCES

Cowdery, J. R. (1984). A fresh look at the concept of tune family. *Ethnomusicology, 28(3)*，495-504.

Deliège, I. (2007). Similarity relations in listening to music: How do they come into play?. *Musicae Scientiae, 11,* 9-37.

Feng, G.Y. (1998). *Chinese Tongzong Folk Songs*. Beijing:China federation of literary and art circles publishing house.

Klusen, E., Moog, H., & Piel, W. (1978). Experimente zur Mündlichen Tradition von Melodien. *Jahrbuch für Volksliedforschung, 2,* 11-32.

Selfridge-Field, E. (2007). Social dimensions of melodic identity, cognition, and association. *Musicae Scientiae,* Discussion Forum 4A, 77-97.

Volk, A., & van Kranenburg, P. (2012). Melodic similarity among folk songs: An annotation study on similarity-based categorization in music. *Musicae Scientiae, 16,* 317-328.

Xu, Y.Y. (2001). My understanding of Tongzong folk songs -- discussion with Prof. Feng. *Huangzhong-Journal of Wuhan Music Conservatory, 3,* 64-69.

Yang, L.L. (2014). *A Brief History of Moo-Lee-Chwa in Modern Times.* Dissertation, Northeast Normal University.

Yang, R.Q. (2002). *Melodic Form of Chinese Folk Songs*. Shanghai: Shanghai Music Publishing House.

reNotate: The Crowdsourcing and Gamification of Symbolic Music Encoding

Ben Taylor[1,2], Matthew Wolf[1], David John Baker[2], Jesse Allison[1,2], Daniel Shanahan[2]
[1]Center for Computation and Technology, Louisiana State University, Baton Rouge, Louisiana, USA
[2]Music Department, Louisiana State University, Louisiana State University, Baton Rouge, USA

Despite advances in music information retrieval, extracting symbolic notation from audio files is still problematic. Currently options for encoding symbolic notation entail either manually encoding scores (which is time-consuming) or making use of optical music recognition software, which requires a human to correct for errors, and is limited to repertoires in which Western musical notation is available. Similar struggles in digitization with print media have been remedied using systems such as "Captcha" and "reCaptcha", which crowdsource the tasks of validation. This paper introduces a similar approach that both crowdsources and gamifies the process of symbolic music encoding.

This project provides a paradigm that focuses on the transcription of non-traditional repertoire that would also be difficult for MIR-related technologies to properly capture. The platform encodes audio data into symbolic notation by transforming the encoding process into a game. Users provide parameters such as tempo, onsets, contour, and eventually melody, and are rewarded for accuracy within a framework that encourages accurate identification of musical material. Data are identified over the course of multiple stages. Users play levels of the game specific to identifying tempo, onset of melodic material, and melodic contour. Once a reliable tempo is determined for piece by enough uses, participants can then tap onsets to allow for the calculation of rhythmic values. A third level asks participants to identify the appropriate contour of a three-note musical event, and a final level implements a probe-tone paradigm to ask the listeners the "goodness-of-fit" of selected pitches within the appropriate contour. This project is able to accurately identify tempi and onsets with a high degree of success, allowing for the rhythmic values of melodic lines to be encoded. These first two levels are available as both a web-service and mobile application.

In the Blink of an Ear: A Critical Review of Very Short Musical Elements (Plinks)

Felix C. Thiesen,[*1] Reinhard Kopiez,[*2] Christoph Reuter,[#3] Isabella Czedik-Eysenberg,[#4] Kathrin Schlemmer [*5]

[*]*Hanover Music Lab, Hanover University of Music, Drama and Media, Hanover, Germany,*
[#]*University of Vienna, Austria,*
[*]*Catholic University of Eichstätt-Ingolstadt, Germany*

[1]felix.thiesen@hmtm-hannover.de, [2]reinhard.kopiez@hmtm-hannover.de,
[3]christoph.reuter@univie.ac.at, [4]isabella@czedik.net, [5]kathrin.schlemmer@ku.de

ABSTRACT

At the turn of the millennium, several much-noticed studies on the rapid assessment of musical genres by very short excerpts ("plinks") have been published. While at first the research mostly focused on clips of popular pieces of music, the recognition of elements in the millisecond range has since been attracting growing attention in general. Although previous studies point to surprisingly short durations of exposure for above-chance genre recognition (mostly between 125 and 250 ms), interpretations of the results are ambiguous due to blurring factors. For example, varying duration thresholds for genre recognition could have been caused by coarse-meshed iterations, unstable audio-quality of sources used, unsatisfactory sample sizes, and insufficient theoretical foundation. In a pilot study, based on a selection of original stimuli used in related precedent studies, we aim to (a) identify essential spectral low-level audio features of the stimuli; (b) determine spectral cues which are obsolete for rapid assessement by means of a correlation analysis between audio features and recognition rates; (c) test for the reliability of genre recognition performance; and (d) compare the selected excerpt to other sections of the particular song (spectral matching). Spectral analyses of audio features are based on the MIR toolbox and on the software package dBSONIC. Replication of genre recognition is based on online-experiments, and the comparison between sound excerpts and the full song is being conducted by a researcher-developed software application. The work is currently under progress but first results indicate that recognition rate strongly depends on the more or less random selection of excerpts in previous studies, which might not always be representative of an entire song. In future studies, researchers should exert more control over the sound material: for instance, by selecting multiple excerpts from one song which should be controlled for the spectral fit to the whole piece.

I. INTRODUCTION

Gray (1942) carried out the first empirical study on minimal durational values required for perception. He limited tape recordings of phonetic vocals to varying short extracts of 3 to 520 ms. Gray observed that even the shortest fragments of acoustic information could be recognized correctly at above-chance levels. He noted that recordings of phonetic vocals with low fundamental frequencies could be reliably detected by 75% of the subjects when only one oscillation of f_0 had been presented to them. For higher-frequency sounds, performance levels of the subjects rose from 30% for a duration of one to 65% at four entire oscillations of f_0.

Referring to Gladwell's (2007) popular scientific bestseller, *Blink: The Power Of Thinking Without Thinking*, Krumhansl (2010) introduced the term "plink" as a very short acoustic stimulus ("thin slice of musical information"). According to her article "Plink: 'Thin Slices' of Music," stimuli in the millisecond range bear potential in the investigation of acoustic cues that are crucial for a rapid and intuitive classification of musical information. Here she referred to a ground-breaking publication by Gjerdingen and Perrott (2008). 400 short stimuli (250, 325, 400, 475, 3000 ms) were derived from prototypical popular compositions representing 10 individual genres (Blues, Classical, Country, Dance, Jazz, Latin, Pop, R&B, Rap and Rock). While the 3-seconds-stimuli were used for the investigation of ceiling effects, the remaining 320 shorter stimuli in a randomized manner. For all of the shorter stimuli, the experimenters reported above-chance recognition levels – increasing with ascending stimulus durations.

In the same year, Schellenberg, Iverson and McKinnon (1999) circumvented the difficulty of unclear genre distinctions by focusing on recognition rates of specific compositions. By asking their participants to name the titles and interpreters of popular songs, they found absolute recognition rates for stimuli of 100 ms and 200 ms. Rentfrow and Gosling (2003) tried to avoid unclear categories by a four-factor model of preference (reflective and complex; intense and rebellious; upbeat and conventional; energetic and rhythmic) which they verified with an extensive sample of n = 1704. A similar approach was followed by Bigand et al. (2005a, 2005b) who tested emotional qualities of short musical excerpts in a model of multidimensional scaling. The different approaches found in these studies, however, do not allow for consistent statements about the processes of recognition on a chronological basis.

II. THEORY

A. Lack of Objectivity in Stimulus Creation

One reason for the variety of minimal durational values proposed for rapid genre assessment (as depicted in Figure 1 on the following page) can be found in stimulus creation processes. None of the studies reviewed for this overview used a standardized process for gaining stimulus materials. In contrary, most of the short excerpts seem to have been gained using unclear criteria. This prevents any statements on how representative the excerpts are for the construct investigated – be it genre, instrumentation or any other perceptible criterion.

Another blurring factor lies in the quality of the source material used for stimulus extraction. Lossy audio compression codecs lead to artefacts that are impossible to control. Some studies do not include any description of the production processes applied – adding to the impression that many blurring factors have mostly been ignored (such as missing "fade-in and -out" which leads to hearable "crackles"). By controlling for these avoidable pitfalls, a higher grade of accuracy in perceived stimulus information can be achieved.

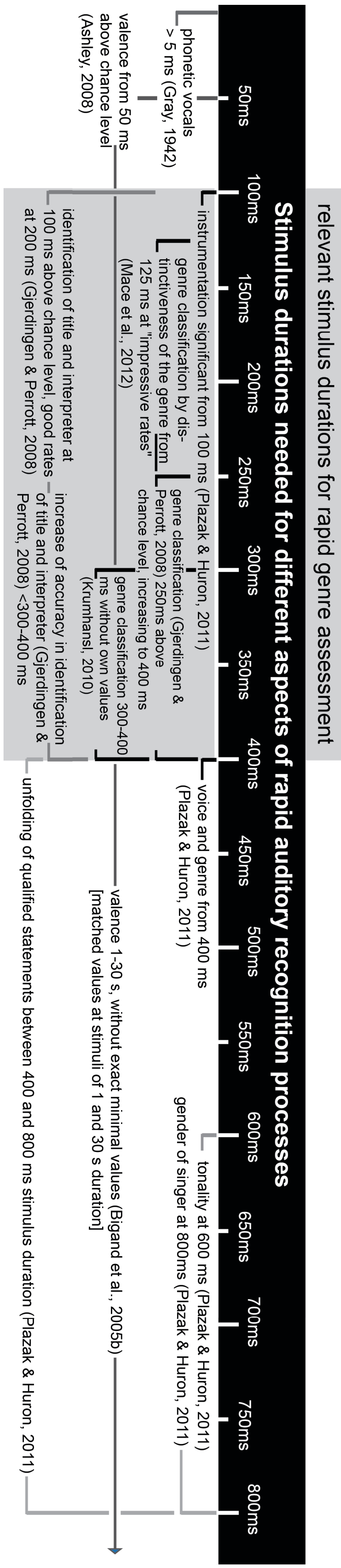


Figure 1. Overview of minimal values proposed in topic-related publications

B. Shortcomings in Existing Literature

Despite the multitude of approaches towards rapid assessment, there is a lack of theoretical embedding: While differing values are reported in literature, there have been few conclusions drawn between these findings and specific psycho-acoustic qualities. Gjerdingen and Perrott (2008) suggested that "timbre might be a factor in the […] classification of instrumental versus vocal excerpts." As of now, there has been no investigation undertaken on what specific aspects of measurable timbre features separate genre spaces. By introducing timbral texture features as referential values, this gap could be closed. This is particularly promising because we can learn from existing models in linguistics and speech recognition technology.

III. METHOD

The shortcomings in research on rapid assessment processes can be tackled by a strict standardization of stimulus generation. We reviewed the stimuli of Krumhansl's 2010 publication. As the original stimuli used in this study were provided as Mp3-files, our first aim was to reconstruct the uncompressed sources used for stimulus generation. Especially for the analysis of timbral texture features, the uncompressed original titles are obligatory to achieve the highest possible accuracy. This necessitated the recovery of the full-quality source-material. 23 out of 28 original albums were found in second-hand CD-shops in northern Germany. As many successful albums (especially those dating back to the 1960's) have been remastered in different versions throughout the years, we decided to use only CDs with (re)release-dates earlier than the date of Krumhansl's publication. Those CDs advertised with extensive remasterings were excluded from the selection. Based on this reconstruction, we were able to use the uncompressed audio signals for the following steps of data analysis.

A. Stimulus Generation Process

In order to avoid subjective confoundations of stimulus selection, a researcher-based script for CSound QT was developed for the automized "plink" production (see Figure 2). This small tool allows the extraction of short stimuli from a pre-defined durational area of a sound-file on the basis of a simple randomization function.

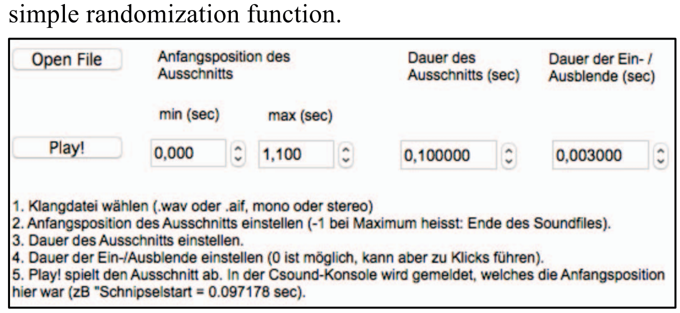


Figure 2. German GUI of the author-based script for randomized "plink" extraction

A beforehand structure analysis of the sound files (subdivision into verses, choruses, etc.) allowed us to set time ranges from which the stimuli should be extracted. To avoid masking noises, we defined fade-in and -out durations before starting the file extraction.

For the first in a row of studies, we designed the "Matryoshka-Method" of stimulus generation (see Figure 3). This algorithm describes the creation of stimuli with descending durations, derived from longer files: A first "plink" with a duration of 800 ms was created. For the extraction of a following 400 ms "plink", the extraction range was set to 3-797 ms of the first one – excluding fade-in and -out areas. Stimuli of 800, 400, 200, 100 and 50 ms were generated to uncover suspected ceiling- and floor-effects of stimulus durations on recognition rates alike. Figure 3 depicts a flow chart diagram of the stimulus production process.

B. Evaluation of Stimulus Content

In a second step of standardization, experts rated the musical content which was perceptible in the short audio files generated – namely, the presence or absence of voice (and therefore gender) as well as the use of percussive and melodic instruments. As the experimental design was to include an online study with predefined answer alternatives, this formed an important reference for the statements gained from the participants. In addition, we offered a free text field for recording additional recognitions of specific genre, interpreter and/or title data. To our knowledge, this is the first study investigating concrete rapid recognition rates of previously determined criteria in terms of musical events.

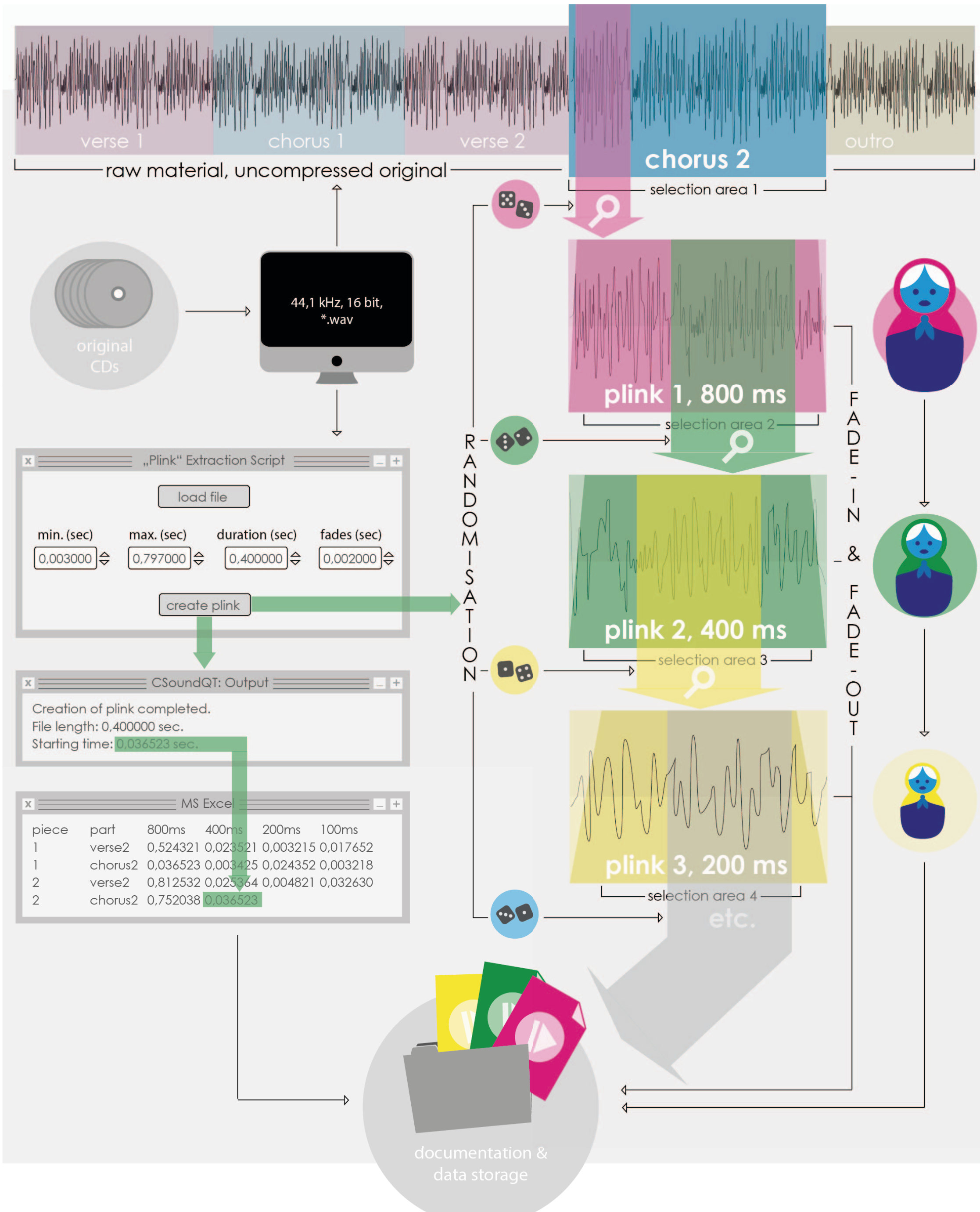


Figure 3. Flowchart of the "Matryoshka" principle of plink production.

C. Timbre Feature Extraction

There are several software solutions available for the extraction of timbral texture features. Among these are freely available packages, such as the "MIR_Toolbox" (Lartillot and Toiviainen, 2007) and the "Genesis Loudness Toolbox" (2007) for *MathWorks' MATLab* as well as high-priced analysis suites like the discontinued software solution *dBSONIC* (2012). What these applications have in common is their ability to extract psycho-acoustic parameters, such as loudness, roughness, and cepstral centroids, from sound files. Accompanying the online data collection of subjective statements, these analyses can contribute to new answers on how subjective recognition rates of specific musical elements correlate with objective psychoacoustic features. Baniya, Ghimire and Lee (2014) proposed a number of timbral texture features that seem promising for genre classification. For instance, Mel-frequency Cepstral Coefficients (MFCCs are generated by converting an audio signal to frames. After a discrete Fourier transformation, a log of the amplitude spectrum is taken. Subsequently, Mel-scaling and smoothing are calculated and a discrete cosine transformation is used to generate the final coefficients). Logan (2000) describes MFCCs as "the dominant feature used for speech recognition."

Along with MFCCs, Baniya et al. suggested *spectral energy, spectral centroid, spectral flux, zero crossing, spectral flatness* and *spectral rolloff* as possible discriminants for genre classification. In our study, a large number of additional values were calculated in order to find those "marker-features" that are suited best for describing rapid assessment processes.

IV. RESULTS

Based on the randomization of stimulus extraction, we expect new and more reliable findings on the minimum duration necessary for correct rapid genre assessment. By reviewing the gained stimuli in an expert survey, we have supplemented the online study design with a retest-reliability reference.

At the present time there are no psycho-acoustic variables linked to specific rapid assessment data. The use of timbral texture features as described in this paper seems very promising for a reinvestigation of this highly relevant topic.

V. CONCLUSIONS

A review of existing literature in the field of rapid genre assessment unveiled several shortcomings in stimulus creation. The introduction of new approaches towards stimulus generation, expert evaluation and timbral feature extraction is considered to shed new light on a research topic that, up until now, has solely been superficially examined. The operationalization of further findings could be relevant for fundamental research on music perception, computer-aided genre classification as well as the improvement of algorithms for hearing-aids.

REFERENCES

Ashley, R., (2008, August). Affective and perceptual responses to very brief musical stimuli. *Proceedings of the International Conference on Music Perception and Cognition, Sapporo, Japan,* 25-29.

Baniya, B. K., Ghimire, D. & Lee, J. (2014, May). Automatic music genre classification using timbral texture and rhythmic content features. *ICACT Transactions on Advanced Communications Technology (TACT), 3*(3), 2014, 434-443.

Bigand, E., Filipic, S. & Lalitte, P. (2005b, December). The time course of emotional responses to music. *Annals of the New York Academy of Sciences, 1060*(1), 429-437.

Bigand, E., Vieillard, S., Madurell, F., Marozeau, J. & Dacquet, A. (2005b, September). Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts. *Cognition & Emotion, 19*(8), 1113-1139.

dBSONIC [Computer software]. (2012). (Version dBSONIC 4.501). (2012) [Computer software for psychoacoustic analysis. For description see http://01db.acoemgroup.com/catalog/Software-1-0-221]. Limonest, France: 01dB-Metravib.

Genesis, Loudness Toolbox [Computer Software], (2007). (Retrieved from http://genesis-acoustics.com/en/loudness_online-32.html)

Gjerdingen, R. O. & Perrott, D. (2008, November). Scanning the dial: The rapid recognition of music genres. *Journal of New Music Research, 37*(2), 2008, 93-100.

Gladwell, M. (2007). *Blink: The power of thinking without thinking*. New York: Back Bay Books.

Gray, G. W. (1942, December). Phonemic microtomy: The minimum duration of perceptible speech sounds. *Communication Monographs 9*(1), 75-90.

Krumhansl, C. (2010, June). Plink: "Thin slices" of music. *Music Perception: An Interdisciplinary Journal, 27*(5), 337-354.

Lartillot, O. & Toiviainen, P. (2007) A Matlab toolbox for musical feature extraction from audio. *International Conference on Digital Audio Effects*, Bordeaux.

Logan, B. (2000, October). Mel frequency cepstral coefficients for music modeling. *Proceedings of the International Symposium on Music Information Retrieval (ISMIR), Plymouth, USA,* (Retrieved from http://ismir2000.ismir.net/papers/logan_paper.pdf)

Mace, S. T., Wagoner, C. L., Teachout, D. J. & Hodges, D. A. (2012, January). Genre identification of very brief musical excerpts. *Psychology of Music, 40*(1), 112-128.

Plazak, J. & Huron, D. (2011, March). The first three seconds: Listener knowledge gained from brief musical excerpts. *Musicae Scientiae, 15*(1), 29-44.

Rentfrow, P. J. & Gosling, S. D. (2003, June). The do re mi's of everyday life: The structure and personality correlates of music preferences. *Journal of Personality and Social Psychology, 84*(6), 1236-1256.

Schellenberg, E. G., Iverson, P. & McKinnon, M. C. (1999, December). Name that tune: Identifying popular recordings from brief excerpts. *Psychonomic Bulletin & Review 6*(4), 641-646.

The Use of Prototype Theory for Understanding the Perception and Concept Formation of Musical Styles

Florian Hantschel,[*1] Claudia Bullerjahn [*2]

[*]*Institut für Musikwissenschaft und Musikpädagogik, Justus-Liebig-Universität, Giessen, Hesse, Germany*
[1]florian.hantschel@musik.uni-giessen.de, [2]claudia.bullerjahn@musik.uni-giessen.de

ABSTRACT

The aim of our study is to investigate the perception of specific musical styles applying the well-established prototype theory of cognitive psychology. Moreover, we wish to demonstrate that the so-called prototypicality of sounding objects directs the inter-subjective decision processes and categorization. Our elicitation (spring, 2015) is based on (1) a self-developed, Lime-Survey-based, online-questionnaire (n_1=764), which consists of typicality ratings for 20 audio-stimuli (also measuring collative, aesthetic-evaluative dimensions & descriptive items), and, (2) a pre-test expert-rating (n_2=4). Audio test-stimuli were taken from previously existing songs, and musical parameters were examined with methods of popular music analysis to deliver supplementary evidence. Questionnaire participants were grouped into six different expertise levels via self-assessment. The statistical analysis of typicality-ratings for n_1 and n_2 clearly indicate that typicality effects control the ratings of the 20 audio-test-stimuli: (a) substantial inter-rater-reliability for typical and atypical items, (b) identification of prototypical items is expertise-independent and (c) relation between stylistically relevant music-parameters and typicality-rating. To conclude, most recent investigations concerning the perception of musical styles either use the linguistic approach of defining specific grammatical rules exclusively, or analyze non-musical parameters, e.g. musicians' image or fan culture. Instead, we argue that psychology's research paradigm of cognitive prototypes provides valuable information for understanding the concept formation and perception processes for the differentiation of musical styles.

I. INTRODUCTION

In modern times, categories and categorization form an integral part of everyday cultural life as systems for the organization of popular culture (cf. Frith 1996, p. 75). Most often we use such aforementioned labels unknowingly. The awareness of their existence and importance often only occurs to us if we encounter situations where we are searching for, but fail to find certain objects. This is because we cannot describe or categorize them with our familiar cognitive concepts. Also, the organization of music into categories such as »genre« or »style« is a Western European phenomenon, which has a long standing cultural history and still has social significance today (cf. Seidel, 1996). To this day, from a music psychological perspective, the cognitive processes involved in the perception and concept-formation of musical styles are still not fully understood. Moreover, research concerning this topic is very sparse.

In the early 1990's, German music psychologist Rainer Niketta[1] studied the value of cognitive prototype theory for understanding the aesthetic perception of musical styles. Since then, to our knowledge, no further quantitative empirical research has been conducted concerning the hypothesis-testing and the development of this worthwhile inquiry.[2] Therefore, the aim of our exploratory, empirical quantitative study is to investigate the perception of specific music styles using the well-established so called prototype theory. This psychological theory is in the tradition of Eleanor Rosch's research from the 1970s onwards on natural objects and language. Moreover, we wish to demonstrate that the so-called prototypicality of sounding objects directs the intersubjective decision processes and categorization.

II. THEORY

A. Theory of music style und conceptual history

The term *style* derives from the Latin word »stilus« originally meaning a writing utensil or a manner of speaking. From the 16th century onwards, the term style was included in music-theoretical treatises (cf. Seidel 1996). Style theories generally pertain to the characteristics of a certain art movement, a group of artworks, an oeuvre/idiolect or a single piece of art. The standard form of style theory is classification, while their operant methods are the acquisition of constant stylistic aspects and the comparison with other styles (cf. ibid.). Because *music style* and *theories of music style* are subject to historic change, the validity and benefit of individual definitions must first be considered. We chose to base our study on Leonard B. Meyer's (1989) general definition:

»[Style is] replication of patterning whether in human behavior or in the artifacts produced by human behavior, that results from a series of choices made within some set of constraints (ibid., p 3).«

B. Music style and genre

In order to theoretically and terminologically distinguish style- and genre-concepts, we define a *music style* sensuo stricto as superordinate category comprised of purely inner-musical elements (cf. Elflein, 2010, p. 31; Moore, 2012, p. 119). Following Peter Wicke's approach (1997, p. 510), those elements include characteristic structural (harmonic, rhythmic, melodic & formal) as well as interpretatorical features (playing techniques, agogic, sound & timbre). As a result of overarching design principles, these elements are common to a group of music pieces.

By contrast, in popular music *genre* commonly includes non-music features as well (e.g. visual image). Unlike in popular-music-discourse, in classical music and traditional German musicology the term »genre« translates into the so called »Gattung«. In his broad analysis of classification, style and genre in musicology Wolfgang Marx (2004, p. 46-58) concludes, that *genre* and *style* are often mistakenly used synonymous. To be correct, he states that the former classification is solely based on form and instrumentation whereas the latter is more complex as it generally describes the way in which something is played (cf. ibid.). Following this interpretation, *genre* can be understood as a sub-category or specific case of *style*.

C. Prototype theory

As a theory, the prototype-approach belongs to the cluster of theories that focus on pattern recognition. These are not congruent with the so-called gestalt theory. In principal, the fundamentals of prototype theory are a result of the linguistically and psychologically orientated studies of Eleanor Rosch (1975, 1978 & 1981). Rosch et al. initially investigated cognitive categorical representation of so-called »natural objects« (colors, forms) and »categories of language-usage in everyday life« (animals or furniture). Subsequently, this research paradigm was transferred to social and human science. Within the scope of empirical psychological research on aesthetics, cognitive music psychology and music theory have adapted the prototype theory to investigate conceptual models and categorization in human thought.

Helga de la Motte-Haber (2005, p. 63) theorizes that a prototype is a form of cognitive typification. Underlying characteristics occurring in various examples are abstracted and determine the shape of a prototype. Thus, they are acquired through experience and learning. This learning process can happen incidentally, e.g. while listening to music. Moreover, according to Niketta (1990), the general prototype-approach postulates that

- boundaries between those cognitive categories formed through abstraction are fluent
- members of categories can vary in their typicality, i.e. how representative they are for that certain category
- for every category various contrastive categories exist
- the ideal representative of a category is called the *Prototype*

However, Gregory L. Murphy (2004, p. 41) states that prototype theory uses *Prototypes* not only in the sense of »best examples« for categories. More recently, the research-paradigm also includes the assumption of prototypes as »summary representations« of concepts. These are portrayable through weighted feature-lists and schemata. Furthermore, German musicologist Martin Rohrmeier (2015, p. 290) states that:

»From a formal perspective, schema or prototype theories and grammar approaches may be construed as the opposite ends of a complexity spectrum. Their primary difference involves a tradeoff in terms of compression: If a pattern evinces both regularity and combinatorial freedom, grammars will be more suitable to describe it; however, if a musical structure exhibits more standardization and less variability, schema-theoretical (and exemplar-based) approaches will be more appropriate.«

The benefit of mentally-represented prototypes is seen in fast reaction times, less erroneous classification and less oblivion in the recognition of patterns (la Motte-Haber, 2005, 63). One of the relevant taxonomic processes involved in cognitive categorization and prototype formation is the »efficiency principle«. This minimizes the number of categories, while at the same time guaranteeing a maximum of informative content (cf. Thorau, 2010, 218). Through the »prototype effect«, new category-candidates are cognitive-statistically judged in relation to existing features in the current category (cf. ibid.).

In his model for the schematic processing of musical structures, Stoffer (1985) already implied that types of style- and form-schemata could be part of a global cognitive music schema. He thereby implicitly assumed that music styles could display a prototypical structure. Martindale and Moore (1989) proclaimed that the perception of music is based on categorizations of mental schemata and prototypicality influences preference judgements (cf. Cohrdes et al. 2011). Subsequently, based on the perceived prototypicality of rock and jazz music Niketta (1990) was the first music psychologist who investigated the correlation between a certain music-style-concept, collative traits and aesthetic-evaluative judgements in an experimental setting.

Because of the idea of pattern recognition, Meyer's musicological theory (1989) is closely related to the cognitive prototype-concept Niketta applied. This connection provides the basis for joining the knowledge acquired through music analysis with prototype-theory for the purpose of understanding the perception and concept formation of musical styles in a multi-methodological approach.

III. METHOD

The musicological groundwork, i.e. a preliminary study combined with a research literature review, resulted in a large corpus of research relevant black-metal-music[3]. The musical parameters were examined qualitatively and quantitatively using methods of popular music analysis and statistics for identifying the relevant stylistic aspects of the music.

Our exploratory, quasi-experimental elicitation is based on a self-developed, Lime-Survey online-questionnaire (n_1=764[4] and a pre-test expert rating (n_2=4). It was published by the server host of Justus-Liebig-University, Giessen, Germany.[5] A promotional campaign was carried out using Facebook, notices, circular email, a newsletter and personal/face-to-face contact. Data was collected April through May 2015. As a further motivation, participants were presented the prospect to participating in a lottery and winning a 50€ coupon for concert tickets.

The main section of the online survey consists of five-point Likert scale-based typicality ratings for twenty audio-stimuli. Each stimulus lasts twelve seconds (incl. fade-in & out).[6] Audio-stimuli were presented one at the time through a Soundcloud-based embedded player device. While filling in the ratings, participants were able to listen multiple times. Once they had rated a stimulus and continued to the next one, moving backward or changing the previous rating was no longer possible.

The twenty audio test-stimuli[7] were taken from an item pool (lexical approach/lexical hypothesis) of 84 existing black-metal-songs. The stimuli were selected in accordance to or in deviation from the stylistically relevant parameters in the pre-test. Afterwards, in a double-blind experimental setting, a typicality rating of four expert[8] for the black-metal-music-style was conducted to secure the fitness of the selected test-items. Another reason for this rating was the subsequent comparison with the ratings of the survey applicants.

The questionnaire-participants were grouped into six different expertise levels by means of a self-assessment question regarding black-metal-music. Applicants with no knowledge of this music were excluded from the survey:[9]

Table 1. Sample description

Group	n	♀	♂	M	SD
1	388	218	163	25.25	6.45
2	140	51	87	27.45	9.63
3	98	3	64	28.01	7.63
4	51	4	47	26.37	5.16
5	54	9	45	27.25	6.85
6	32	1	32	27.94	4.63

The statistical analysis was carried out using SPSS/PASW 18 statistics. Multivariate statistics were used to determine the effects of descriptive variables on the ratings in the main section of the survey. Revising Niketta's approach (1990), seven prototypicality-scales (overall group-average & six for each expertise-group) consisting of mean ratings for the twenty audio-stimuli were calculated to measure inter-correlations between the different expertise levels. Regression analysis was used to measure the relationship between the perceived prototypicality and judgement consistency of the participants (operationalized through the use of standard deviation from the 20 mean typicality-ratings). Furthermore, the absolute frequencies of the typicality ratings for every stimulus were calculated, and a k-cluster analysis (k=5) was executed to compare the results of the expert-ratings with those of the survey participants. The inter-rater reliability for the expert rating and the ratings of the participants were calculated using the Fleiss'-Kappa method. Non-parametric tests were used for measuring possible expertise-differences in the typicality-ratings of the twenty audio stimuli (non-parametric, because the expert groups were not of equal size). The Kruskal-Wallis-test was used to determine a possible relationship between musical parameters and typicality-ratings. For this purpose, audio-test-stimuli were repeatedly grouped according to stylistically relevant parameters (e.g. vocals vs. without vocals) and statistically compared.

IV. RESULTS

A. Pre-test corpus-analysis and research literature review[10]

The classic heavy metal band instruments (in the tradition of rock bands) plus optional use of keyboard instruments are generally combined with a predominantly male, highly throaty-croaking and screaming/screeching voice. Clean-singing is only used sporadically. The strings of the electric guitars are often played unmuted with so-called fast tremolo-pickings in monophonic or polyphonic modal-interval- or harmony-texture. Broken guitar-chords with single-note-picking are also commonly used. When combined with these guitar playing techniques, the high density of dissonance, the heavy distortion (inharmonic overtones) and the common frequency-modulation – featuring lots of reverb and treble-equalization – result in a droning soundscape. Guitar-solos are usually interspersed. The electric bass often only has a pseudo-presence in the band's context, because of the attenuation of bass frequencies. Furthermore, the so-called blast-beats with a maximum of impulse density are style-defining for drums and percussion. Also variations of a reduced back-beat – with or without double-bass playing – are used.

In contrast to previous analyses, statistically we could not identify a characteristic tempo of black-metal-music. Moreover, rhythms in general have a minor relevance. A very characteristic, tonal-harmonic feature of black-metal-music is the organization of riffs, phrases and large-form in so called tone-fields (Cohn, 2004; Haas, 2004). Within those fields, black-metal-music predominantly or even exclusively uses minor harmonies/chords and minor mediants. Still, the so called power-chord [11] often forms an integral part of black-metal-guitar-riffing. Overall, especially important to black-metal-music is an atmospherically blurred sound-aesthetic with a lack of transparency and audibility (low fidelity). Combined with a so-called »trueness«-ideology it serves as an authenticity construction.

B. Expert rating

The Fleiss' Kappa calculating the inter-rater reliability for the twenty audio-test-stimuli broadly ranges between values 0.17 – 1 (low to perfect agreement). On the contrary, the mean Fleiss' Kappa of 0.76 suggests an overall substantial agreement of all four black-metal-music-experts. The quick cluster analysis with forced distribution (k=5) revealed five cluster-centers (C1 = 3; C2 = 4; C3 = 4; C4 = 5 & C5 = 4). The statistical analysis of variance (ANOVA) displays a highly significant ($p<0.01$) main effect, which means there is maximum distance between the items in the five clusters. The results of the expert-rating and additional information can be found in appendix A. In an accompanying interview, all four experts approved of the representability of the twenty audio-test-stimuli and offered their thoughts on the appropriateness of the statistical grouping.

C. Online survey

All in all, the descriptive variables of respondents (e.g. age, professional qualification, musician's expertise, media expertise) have no detectable multivariate (MANOVA) influence on the rating of the music samples. Only a general verbal preference for metal-music has statistical significant impact on the ratings in the main section of the online survey. However, this relationship between metal-preference and typicality-rating are confounded by black-metal-expertise-group. Therefore, when partialled out, the effect disappears. Compared to Niketta's study, the statistically-formed average typicality-rating for each of the twenty audio-test-stimuli proves to be a non-representative measure for determining relationships between single variables of judgment (typicality, collative, aesthetic-evaluative). This is, because only through the differentiation in black-metal-expertise-groups we determine significant correlational relationships.

The ratings of the expert-sample and the expertise group 6 are particularly similar (high correlation, significant prediction of rater judgement possible through black metal expertise group 6 via regression, inter-rater-reliability), suggesting that both have a similar level of expertise. Almost all music stimuli are judged quite similarly by participants of all black-metal expertise-groups. An exception is audio-stimulus 7, which was deliberately included as a stylistically closely related control stimulus (death-metal-music). Here, the ratings of expertise-groups 1 and 6 are diametrically opposed and almost all intermediate stages (except 3 and 4) significantly ($p\leq0.001$) differ in their judgment.

Only expertise group 6 can also identify music excerpts without the characteristic vocals (stimulus 6) as typical. Importantly, the audio test stimuli 12, 13, 15 and 20, assessed as being very typical by the four experts, are also on average judged as being typical by all of the six groups from the survey. This means, that the recognition of prototypical black-metal-music is independent from expertise.

Confirmed by non-parametric group comparison of audio test-stimuli, all of stylistically relevant musical parameters detected in the pre-test-analysis (screeching/screaming vocals, tremolo-picking drones and blast-beat drum-patterns) lead to high typicality-ratings. Yet, an even more significant and therefore fundamental criterion of typicality-perception is the aesthetics of lo-fi production. The regression analysis (dependent variable: mean typicality-rating; predictor: standard-deviation) shows ($p \leq 0.01$; $R^2 = 0.670$) that with increasing levels of perceived prototypicality, the expertise groups 1 and 6 independently displayed higher levels of inter-individual consistency of judgement.

On the one hand, with 88% of participants rating it as typical example, the audio test stimulus 20 (Tatthogghua, – Status Stürmer) appears to be the best example and thus can be called the prototype of the study. On the other hand, example 9 (Lantlôs – Minusmensch) with 91% survey contributors classifying it as the most atypical representative, consequently can be called the antitype.

The results of the analysis of free-text answers show that 68.6% of the respondents identify one of nine bands as being a typical black-metal-band, most of which are Norwegian (decreasing order: Darkthrone, Immortal, Mayhem, Dimmu Borgir, Gorgoroth, Behemoth, Marduk, Burzum, Venom). Altogether, higher expertise-groups' answers are more versatile answers. The most frequent free-text answer is the song »Transilvanian Hunger« by the Norwegian band Darkthrone and can thus be called the verbal prototype of the survey.

V. DISCUSSION

All in all, a point of criticism is the quasi-experimental setting of our study. In the future, controlled laboratory experiments should be performed to minimize the possible impact of confounding factors. Further, although the groups 1 to 6 reflect a representative cross-section of German society, these should be approximated to near equal size and gender-distribution to ensure comparability. Then, other statistical methods such as parametric test procedures (e.g. t-test; ANOVA) are possible as well. Another criticism concerning our study is that the expertise for black-metal-music was only measured indirectly. Hence, it can be assumed that the self-assessment assignment was not sufficiently objective. Certainly, the choice of twenty audio-test-stimuli for an exploratory and hypothesis-testing investigation was reasonable. In respect to the small amount of test-stimuli, it is also obvious that there are no controlled results for each individual stylistically-relevant parameter possible. That is why we recommended that selected features could be examined in a controlled manner using a large corpus of test-stimuli and expert-raters. Another possibility is to reflect the holistic processing of music perception and cognition. With this in mind, one might apply the typicality rating-paradigm in an experimental context to a corpus of pre-existing full-song-examples instead of short cues. Under specific instructional conditions (e.g. "Pay special attention to…"), repeated measures would be possible as well. Unlike Niketta (1990) we chose to use no contrast categories and no exclusion of items (e.g. "Is not Black Metal" was no rating option). Future experiments should also implement multipolar ratings or rating scales for various different style-categories.

Different from Rosh et al. (ibid.), the inter-rater-reliabilities concerning sounding aesthetic-objects are, in line with expectations, lower as for »natural objects« and language. This could be due to:

- category vagueness (debates on the so called distinction of first & second wave of black metal compared to proto-black and black metal proper are still relatively new)
- the test subjects generally overestimated their expertise
- overseen confounding variables have systematic or non-systematic impact
- uncertainty about the category, because of its specifity
- forced involvement of language processing and declarative knowledge (both rely on slower and conscious paths of cognition compared to auditory perception of music and/or emotional processing)

Hence, keeping in mind, that prototypes are believed to emerge incidentally as product of implicit learning, it is quite reasonable to assume that the very immediate process of categorization has to be captured wordless (for example with imaging methods, such as EEG or fMRI). Moreover, it might be useful to adjust the particular scale in use (insofar one wants to use rating-scales after all) according to the individual expertise of the test-subject (for professionals a multi-stage measurement and for laymen one as easy and understandable as possible). Following Murphy's review and interpretation (2006), it is quite possible that lay judge more exemplar-based and professionals by means of an internalized prototype. The overall question is how the typicality-rating comes into being and which parameters are influential. The multivariate tests suggest that in the questionnaire, socio-demographic variables and also verbally recorded music-preferences have no statistically significant impact on the typicality-rating. The only variable that comes into question is the membership of a black-metal-expertise group. Clearly, regression-analysis shows that subjects of a higher group affiliation judge most similarly to the expert-raters from the pre-test. Despite the highly significant differences, all in all, only ten of the twenty test-stimuli were rated significantly different and it must be noted that:

- the very typical representatives were evaluated in a comparable manner in each black-metal-expertise-group
- only for example 7 the average ratings are substantially different and group 1 is diametrically opposed to group 6, which suggest an effect of expertise
- e.g. for the prototype-stimulus 20, which, after all , was rated as very typical by all expertise groups, the statistical significance is most likely also be due to the very large sample size

Following the above-mentioned assumptions of prototype-theory, this leads us to the conclusion, that these typicality-effects suggest an inter-subjectively shared underlying concept of what is prototypical black-metal-music.

VI. CONCLUSION

To conclude, most recent investigations concerning the perception of musical styles either use the linguistic approach of defining specific grammatical rules (e.g. Martin Rohrmeier's Generative Syntax Model of tonal harmony, 2005, 2009, 2015), or analyze non-musical parameters, for example musicians' image or fan culture. Instead, we argue that psychology's research-paradigm of cognitive prototypes provides valuable information for understanding the concept formation and perception processes for the differentiation of musical styles. We used the example of black-metal-music, which is an extreme sub-style of heavy metal music. Excluding those participants who never heard of this music before, ratings of twenty audio-test-stimuli revealed, that typicality effects govern the intersubjective perception of what are typical and untypical sound examples for this music. Moreover and very surprising in our quantitative-empirical study with 764 participants from very different backgrounds and expertise, there is substantial agreement concerning the very ends of the typicality continuum. Against all odds that black-metal is a minority's music where only a chosen few are in the secret,. prototypical black-metal-music could be identified expertise-independently. Concerning future research, more experiments following the paradigms established by Rosch et al. (ibid.), especially regarding the very basic meta-categories of everyday social life, such as e.g. rock, classic, jazz or electronic dance music, are preferable.

ACKNOWLEDGMENT

We thank the German National Academic Foundation, the heavy-metal supporters-community *Undergrounded*, Ursula Kruck-Hantschel, Thilo Marauhn, Philipp Dietzel, Joey Sato, Nils Wilken and Lydia Scott for their support.

REFERENCES

Chaker, S. (2011a). Extreme Music for Extreme People: Black and Death Metal. Put to the Test in a Comparative Empirical Study. In: *The Metal Void. First Gatherings*. N. W. Scott & I. v. Helden (Eds.). Oxford: Inter-Disciplinary Press, 265-278.

Chaker, S. (2011b). Götter des Gemetzels. Mediale Inszenierung von Tod und Gewalt im Death Metal. In R. Bartosch (Ed.), *Heavy Metal Studies. Band 1: Lyrics und Intertextualität*. Oberhausen: Verlag Nicole Schmenk.

Chaker, S. (2014). *Schwarzmetall und Todesblei. Über den Umgang mit Musik in den Death- und Black-Metal-Szenen Deutschlands*. Berlin: Archiv der Jugendkulturen.

Cohn, R. (2004). Resemblances: Tonal Signification in the Freudian Age. In *Journal of the American Musicological Society*, 57 (2), 285-324.

Cohrdes, C., Lehmann, M. & Kopiez, R. (2011). Typikalität, Musiker-Image und die Musikbewertung durch Jugendliche. *Musicae Scientae*, 16 (1), 81-101.

Cohrdes, C. & Kopiez, R. (2014). Optimal distinctivness and adolescent music appreciation: Development of music and image-related typicality scales. *Psychology of Music*, 1-18.

Dietzel, P. (2013). *Blizzard Bea(s)ts. Die musikalische Ästhetik des norwegischen Black Metal in der ersten Hälfte der 1990er Jahre.* Magister-thesis. Justus-Liebig-University, Giessen.

Eckes, T. (1991). *Psychologie der Begriffe. Strukturen des Wissens und Prozesse der Kategorisierung*. Göttingen.: Hogrefe.

Elflein, D. (2010). *Schwermetallanalysen. Die musikalische Sprache des Heavy Metal*. Bielefeld: Transcript.

Frith, S. (1996). *Performing Rites. On the Value of Popular Music*. Oxford: Oxford University Press.

Haas, B. (2004). *Die neue Tonalität von Schubert bis Webern. Hören und Analysieren nach Albert Simon*. Wilhelmshaven: Noetzel.

Hagen, R. (2011). Musical Style, Ideology, and Mythology in Norwegian Black Metal. In J. Wallach, H. M. Berger & P. D. Greene (Eds.), *Metal Rules The Globe. Heavy Metal Music Around The World* (pp. 180-199). Durrham & London: Duke University Press,

Hainaut, B. (2012) "Fear and Wonder". Le fantastique sombre et l'harmonie des médiantes, de hollywood au black metal. *Volume! The French Journal of Popular Music Studies*, 9(2) (*=Contre-cultures*), 179-197.

La Motte-Haber, H. de (2004). Ästhetische Erfahrung: Wahrnehmung, Wirkung, Ich-Beteiligung. In *Musikästhetik. Handbuch der systematischen Musikwissenschaft*. H. de la Motte-Haber & E. Tramsen (Ed.). Laaber: Laaber, 408-429.

La Motte-Haber, H. de (2005). Modelle der musikalischen Wahrnehmung. Psychophysik – Gestalt – Invarianten – Mustererkennen – Neuronale Netze – Sprachmetapher. In H. de la Motte-Haber & G. Rötter (Eds.), *Musikpsychologie. Handbuch der systematischen Musikwissenschaft* (pp. 55-73). Laaber: Laaber.

La Motte-Haber, H. de (2010). Wahrnehmung. In H. de la Motte-Haber, H. v. Loesch, G. Rötter & C. Utz (Eds.), *Lexikon der systematischen Musikwissenschaft* (pp. 552-526.). Laaber: Laaber.

Martindale, C. & Moore, K. (1989). Relationship of musical preference to collative, ecological, and psychophysical variables. *Music Perception*, 6(4), 431-445.

Marx, W. (2004). *Klassifikation und Gattungsbegriff in der Musikwissenschaft*. Hildesheim: Georg Olms Verlag.

Meyer, L. B. (1989). *Style and Music: Theory, History, Ideology*. Philadelphia: University of Chicago Press.

Moore, A. F. (2001). *Rock: The Primary Text. Developing a musicology of rock*. Burlington: Ashgate.

Moore, A. F. (2012). *Song Means: Analysing and Interpreting Recorded Popular Song*. Burlington: Ashgate.

Murphy, G. L. (2004). *The Big Book of Concepts*. Cambridge, Massachusetts: MIT Press.

Niketta, R. (1990). Was ist prototypische Rockmusik? Zum Zusammenhang zwischen Prototypikalität, Komplexität und ästhetischem Urteil. In K-E. Behne, G. Kleinen & H. de la Motte-Haber (Eds.), *Musikpsychologie. Empirische Forschungen – Ästhetische Experimente. Jahrbuch der deutschen Gesellschaft für Musikpsychologie* (pp. 35-57). Wilhelmshaven: Florian Noetzel Verlag.

Reyes, I. (2013). Blacker than Death: Recollecting the "Black Turn" in Metal Aesthetics. *Journal of Popular Music Studies*, 25 (2), 240-257.

Rohrmeier, M. (2009). *Implicit learning of musical structure: Experimental and computational approaches*. PhD Dissertation. University of Cambridge.

Rohrmeier, M. & Neuwirth, M. (2015). Towards a Syntax of the Classical Cadence. In M. Neuwirth & P. Bergé (Eds.), *What Is A Cadence. Theoretical and Analytical Perspectives on Cadences in Classical Repertoire* (pp. 287-338). Leuven: Leuven University Press.

Rohrmeier, M. (2005). *Towards modelling movement in music: Analysing properties and dynamic aspects of pc set sequences in Bach's chorales*. MPhil dissertation. University of Cambridge.

Rosch, E. & Mervis, C. B. (1981). Categorization of natural objects. *Annual Review of Psychology*, 32, 89-115.

Rosch, E. & B.B. Lloyd (1978). *Cognition and categorization*. Hillsdale, New Jersey: Erlbaum.

Rosch, E. & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7, 573-605.

Rötter, G. (2005). Musik und Emotion. Musik als psychoakustische Substanz – Musikalischer Ausdruck – Neue experimentelle Ästhetik – Emotionstheorien – Funktionale Musik. In H. de la Motte-Haber & G. Rötter (Eds.), *Musikpsychologie. Handbuch der systematischen Musikwissenschaft* (pp. 268-338). Laaber: Laaber.

Seidel, W.(1996). Stil. In L. Finscher (Ed.), *Die Musik in Geschichte und Gegenwart*. *Sachteil Bd. 8* (pp. 1740-1765). Kassel u.a., 2. Aufl. 1996.

Stoffer, T. H. (1985). Modelle der kognitiven Verarbeitung und Repräsentation musikalischer Strukturen. In O. Neumann (Eds.), *Perspektiven der Kognitionspsychologie* (pp. 147-184). Berlin: Axel Springer Verlag.

Thorau, C. (2010). Kognitive Musiktheorie. In H.de la Motte-Haber, H. v. Loesch, G. Rötter & C. Utz (Eds.), *Lexikon der systematischen Musikwissenschaft* (pp. 217-219.). Laaber: Laaber.

Wicke, P. (1997). Stil. In P. Wicke, K.-E. und W. Ziegenrücker (Eds.), *Handbuch der populären Musik*. Wiesbaden: Atlantis (Schott).

APPENDIX A

Table 2. Cluster-Center-Analysis of Expert-Rating (also including name of band, song title, stimulus ID [1-20], year of release and record label)

Proto-typicality	Band	Song (stimulus ID; year; album; record label)
Very typical	Nocte Obducta	Fick die Muse (13; 2001; *Schwarzmetall*; Grynd Syndicate Media)
	Tsatthoggua	Status Stürmer (20; 1998; *Trans Cunt Whip*; Osmose Productions)
	Nocte Obducta	Hunde des Krieges (12; 2001; *Schwarzmetall*; Grynd Syndicate Media)
	Nyktalgia	Misere Nobis (15; 2004; *Nyktalgia*; Noc Colours)
typical	Imperium Dekadenz	Fields of Silence (8; 2006; *...und die Welt ward karg und leer*; Perverted Taste)
	Ascension	Grey Light Sibling (3; 2010; *Consolamentum*; WTC)
	Membaris	Monotonkrieger (11; 2012; *Entartet*; ARTicaz)
	Eis	Wetterkreuz (6; 2012; *Wetterkreuz*; Prophecy Productions)
	Nocte Obducta	Sequenzen einer Wanderung Teil 2 (14; 2008; *Sequenzen einer Wanderung*; SCR)
typical/ untypical	Aeba	Des Zornes Banner (2; 2001; *Rebellion*; Last Episode)
	Samsas Traum	Schwert deiner Sonne (17; 2007; *Schwert deiner Sonne*; Trisol Music Group)
	ColdWorld	Hymn to Eternal Forest (4; 2008; *Melancholie*; Cold Dimensions)
	Lantlôs	Pulse/Surreal (10; 2010; *.neon*; Lupus Lounge)
untypical	Thulcandra	Ritual of Sight (19; 2011; *Under A Frozen Sun*; Napalm Records)
	Agathodai-mon	Near Dark (1; 1998; *Blacken The Angel*; Nuclear Blast)
	Disbelief	The Thought Product (5; 2007; *Navigator*; Massacre Records)
Very untypical	Obscura	Vortex Omnivium (16; 2011; *Omnivium*; Relapse Records)
	Secrets of the Moon	Seven Bells (18; 2012; *Seven Bells*; Lupus Lounge)
	Fleshcrawl	Made of Flesh (7; 2004; *Made of Flesh*; Metal Balde Records)
	Lantlôs	Minusmensch (9; 2010; *.neon*; Lupus Lounge)

[1] Niketta (1990) investigated in the two major categories »Rock« and »Jazz«.

[2] However, fundamental research approaches concerning prototype-theory and singled out musical parameters such as melody, rhythm or harmony do exist (e.g. Rohrmeier et al., 2015).
Chordes et al. (2011, 2014) did in fact use typicality-ratings, but only focused on the influence of the non-musical parameter »visual image of musician« for categorization of musical stimuli into different *genre*.

[3] All investigations where conducted in the course of Florian Hantschel's master-thesis, which specializes on style and sub-stylistic branching in heavy metal music. The focus lies on black metal music.

[4] 1543 participants started the survey, of whom only 764 were fully filled-in and valid, 576 invalid and 203 dropped by exclusion criterion.

[5] Online access, further information and audio-test-stimuli available via https://opinio.hrz.uni-giessen.de/limesurvey/ index.php/587645?lang=ie.

[6] Collative, aesthetic-evaluative dimensions & descriptive items were measured in a comparable manner. Socio-demographic variables (age, gender, musical background) and music preferences were collected as well as free text naming concerning a typical black metal band and song.

[7] Those are specified in the appendix A and sorted according to the result of the expert-rating.

[8] Male, musicians, min. 10 years of expertise

[9] Shrinking group size was to be expected, because the groups were formed on the basis of black metal expertise level and therefore represent a social cross-section. Male dominance can easily be explained by the target audience of this study, which are primarily heavy metal attracted persons (mean preference for metal music in this sample M=3,57 whereas 5 was the highest possible rating). According to a broad quantitative-empirical surveys of musicologist Sarah Chaker (2011a, 2011b & 2014), German extreme metal scene is male dominated by round about 85 percent.

[10] Black metal music centered literature for this review was Dietzel 2013, Hagen 2012, Hainaut 2012 and Reyes 2013.

[11] Genre-defining in rock and especially heavy-meta-music

Exploring circadian patterns in musical imagery

Freya Bailes[1]
[1]School of Music, University of Leeds, Leeds, UK

Musical imagery is the conscious experience of an internal representation of music; a form of inner hearing. Experience sampling methods (ESM) have been used to study musical imagery as it occurs in everyday life, but findings have been inconsistent regarding its frequency or strength as a function of the time of day at which it is experienced (Bailes, 2015; Floridou & Müllensiefen, 2015). Unpublished analyses of ESM data from Bailes (2015) will be presented to show no relationship between the time of day and imagery vividness [$F(4, 179) = 0.09$, $p > .05$]. However, the frequency of musical imagery over the course of the day approximates the human circadian rhythm, rising to a peak at midday and falling into evening. Experimental research into the chronobiology of musical imagery missing, but there is evidence consistent with circadian patterns in the vividness of visual, motor, and auditory imagery (Gueugneau & Papaxanthis, 2010; Hubbard, 2010; Kokoszka, Domoslwski, Wallace, & Borzym, 2000). This poster asks whether there is a *theoretical* basis to experimentally test for circadian patterns in musical imagery. Its aim is to set out a theoretical framework for future work. The method identifies and examines common features of musical imagery and forms of spontaneous cognition such as intrusive thoughts and mind wandering (e.g. Floridou & Müllensiefen, 2015; Hyman et al., 2013). The pertinence of theories from general psychology such as the 'Basic Rest Activity Cycle (BRAC)' (Duchniewska & Kokoszka, 2003), and the chronobiology of Rossi & Rossi's (2008) four-stage creative cycle is also assessed. The work results in a theoretical framework that can be used to derive testable hypotheses. It also raises questions about the potential for our biology to influence musicians' voluntary uses of musical imagery. With reasonable grounds to do so, composers and performers might maximise their creative imagery through an awareness of its relationship to their circadian cycle.

Musical factors influencing tempo determination

Leigh VanHandel[1]
[1] College of Music, Michigan State University, East Lansing, MI, USA

Research on tempo has concentrated on the concepts of *preferred tempo* (i.e., spontaneous tapping rate or preferred rate of activity), *tempo memory* (tempo memory/imagery for familiar songs, or group memory for tempo), and *tempo identification/discrimination* (the ability to identify or discriminate different tempos). However, there has been very little research on *tempo determination,* or the process of determining an appropriate tempo for an unfamiliar musical excerpt.

This study investigates *tempo determination* with an experiment designed to establish which musical characteristics provide contextual cues for appropriate tempo in an unfamiliar musical excerpt. The musical characteristics include metrical/hypermetrical, melodic, and rhythmic characteristics. For example, Boltz (1998) determined that in monophonic melodies, excerpts containing multiple contour changes, more/larger pitch skips, and melodic accents at odds with the meter were judged to be slower than reference excerpts. While Boltz's study was a tempo discrimination task, her results indicate that melodic characteristics have an important effect on the perceived tempo of a musical excerpt, and that these melodic characteristics might play a similar role in a tempo determination task.

In the experiment, subjects are presented with novel musical stimuli designed to study one of the musical characteristics. Subjects manipulate the tempo of the stimuli in real time, with no pitch shifting, via a controller; their task is to determine what they believe is an appropriate tempo for each excerpt through manipulation of the controller. When subjects determine an appropriate tempo, their response is recorded and they move onto the next excerpt.

Through principal component analysis, it is possible to determine which of the musical characteristics are of primary importance in subjects' tempo determinations. Preliminary results indicate that melodic and rhythmic characteristics are of primary importance.

Relative Salience in Polyphonic Music

Marco Bussi,*[1] Sarah Sauvé,#[2] Marcus Pearce #[3]

* *Dipartimento di Musicologia e Beni Culturali, Università degli Studi di Pavia, Italy*

School of Electronic Engineering and Computer Science, Queen Mary University of London, United Kingdom

[1] marco.bussi01@universitadipavia.it, [2]s.a.sauve@qmul.ac.uk, [3]marcus.pearce@qmul.ac.uk

ABSTRACT

Music is often composed of different streams defined by different perceptual parameters. The phenomenon of auditory streaming has been studied in terms of pitch, duration and timbre among others. The present experiment seeks to examine the effect of three musical features (harmony, melody and rhythm) on the segregation of a target melody, to map the relative perceptual salience of these features in a polyphonic context. 32 participants, 16 musicians and 16 non-musicians were asked to report whether a *target melody* presented in a polyphonic context was the same or different from when it was presented alone. Variations of the melody were created by manipulating two durations and two pitches, one early and one late for each. The target melody was "hidden" amongst masking lines that were systematically and independently varied on three levels of melodic, harmonic and rhythmic complexity. A computational analysis of the stimuli was also carried out, where onset and offset synchrony, pitch comodulation and harmonic overlap were computed and used to predict behavioural responses. Though the behavioural results do not show an effect of manipulated complexity on perceptual salience, it was found that when the changes in the target melody were late rather than early, they were significantly easier to perceive. The computational analysis revealed that harmonic overlap alone predicts whether the target will be correctly identified as same or different. Moreover, the results demonstrate the importance of musical training in melodic segregation, showing higher accuracy and lower thresholds in musicians.

I. INTRODUCTION

Music provides complex acoustic signals, deriving from the superposition, in lines or streams, of complex sounds with different internal characteristics. The ability to perceive such streams and to group them according to common characteristics is a primary goal in music listening, especially in polyphonic contexts, where this ability allows recognition of melodies and their division from accompanying musical contexts.

The phenomenon of auditory grouping has been studied through basic principles of Gestalt (Wertheimer, 1923) and in a number of auditory parameters including frequency, intensity, location, periodicity and timbre (Iverson, 1995; van Noorden, 1975; Vliegen, Moore, & Oxenham, 1999; for a complete review see Bregman, 1990). Attention has been increasingly focused on determining the relative importance of perceptual cues in auditory streaming to better understand the perceptual organization of complex sound sequences, such as polyphonic pieces of music (Duane, 2013; Prince, Thompson, & Schmuckler, 2009).

A recent study by Marozeau and colleagues (Marozeau, Innes-Brown, & Blamey, 2013) measured relative effects of frequency, temporal and spectral envelopes on auditory streaming. In this experiment participants rated the difficulty of segregating a four-note melody interleaved with distractor tones. The results showed that when the separation between melody and distractor tones decreased, the segregation of the target melody became more difficult, and that musicians and non-musicians used different cues to achieve streaming, as well as musicians tolerating objectively more difficult listening situations (i.e., greater pitch overlap).

The relation between pitch and time has been investigated by Prince and colleagues (Prince et al., 2009) using a series of behavioral paradigms based on goodness-of-fit ratings of a probe event heard after a tonally and metrically well-defined musical context. The results showed these dimensions to be asymmetrically integrated with a higher importance given to pitch cues, which appear to be much more salient in common-practice Western music, presumably due to higher complexity, as defined by number of different pitches and number of different rhythmic durations.

In another recent study, Duane (Duane, 2013) calculated onset synchrony, offset synchrony, pitch comodulation and harmonic overlap and used these values to model stream segregation results of a behavioural study using the expositions of 18th and 19th century string quartets. These four measures were chosen as likely candidates to predict integration or segregation, based on Gestalt rules and current knowledge of auditory stream analysis (Bregman, 1990). For example, pitch comodulation indexes similarity of movement between two lines. If lines move in parallel thirds, they are likely to be heard as one perceptual stream, whereas if lines are in contrary motion, they are more likely to be segregated.

How all these segregating factors interact is not clear yet and there is no consensus in the literature on whether they should be interpreted as integrated or independent dimensions

The present experiment seeks to examine the effect of three different musical features on the segregation of a target melody: harmony, melody and rhythm. This melody will be masked among distractor melodies which will be independently varied between three levels of complexity on three dimensions: *harmonic progression*, *melodic complexity* and number of *rhythmic note durations*. As salience has been suggested to be related to complexity (Prince et al., 2009), our hypothesis is that increasing the complexity of the masking lines will make them more salient, consequently making the target melody more difficult to segregate. This particular design should allow us to map the relative perceptual salience of melody, rhythm and harmony in a polyphonic context and to analyze how these features interact using both behavioural (Marozeau et al., 2013) and computational (Duane, 2013) methods.

II. MATERIALS AND METHODS

A. Participants

32 participants (10 female, mean age 27.4, SD = 10.2) took part in this experiment: 16 musicians, with at least 7 years of formal training, and 16 non-musicians, with less than 1 year of formal training.

B. Behavioural stimuli

For the purpose of the experiment a series of 7-voice polyphonic stimuli was created. The seven voices are written to cover an entire orchestral range and are played by the following instruments: one bass organ, two bassoons, one Clarinet and three flutes, respectively located in the following subjective registers: very-low, low, medium-low, medium, medium-high, high and very-high.

The stimuli are composed according to the basic rules of harmony and voice leading and according to standard rules of functional harmony, as systematized by Schönberg (Schönberg & Stein, 1969).

In line with the proposition that salience is linked to complexity, stimuli were constructed to be equally complex as a baseline, with 7 different voices, 7 different rhythmic duration values, 7 different pitches and 7 different chord species. The latter three were then systematically and independently manipulated in order to manipulate complexity and evaluate its effect on perceptual salience. Similarly to Marozeau and colleagues (Marozeau et al., 2013), we designed our stimuli with one target (the middle) voice and six distractor (top and bottom three) voices. The target melody, played by the clarinet, is the same in every composition save experimental manipulation, described below, while distractor voices are systematically manipulated in complexity.

The target melody was composed according very strict criteria: it has seven different pitches (C major diatonic scale), seven different rhythmic durations (proportions 1:1; 1:2; 1:3; 1:4; 1:6; 1:8; 1:10; where the shortest value is a sixteenth note) and seven different direct (between adjacent notes) and indirect (between non-adjacent notes) intervals (minor second, major second, minor third, major third, fourth, tritone and fifth). The melody has a well-defined, convex melodic contour, developing over a ninth range with only one high peak reached in its final part. Therefore, the participant should be able to pick out the target melody from among the other voices by recognizing its characteristic contour.

Further to the above specifications, the target melody was manipulated for same/different experimental task purposes. Four changes were possible: one altered note in the middle of the melody, one altered note duration in the middle of the melody, one altered final note or one altered final rhythmic duration. Alterations are illustrated in Figure 1.

The distractor voices are also constructed according to very strict constraints, varying on three parameters: harmonic progression, melodic complexity and number of rhythmic durations, where each parameter varies among three levels of difficulty: simple, medium and complex.

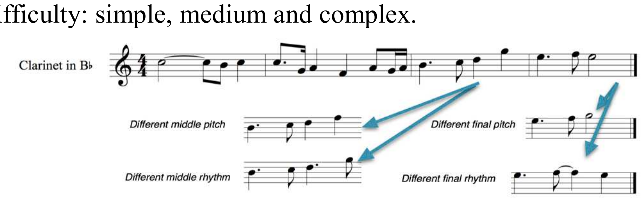


Figure 1. Target melody and four change typologies

The rhythmic durations used in the distractor voices are selected from the same set as in the target melody. The simple level has only two rhythmic durations in each and across all six voices. These two values are the longest rhythmic values possible, in accordance with the frequency of changes in the harmonic progression. In the medium level there are three rhythmic durations in each and across all six voices. In the complex level there are at least four different rhythmic durations in each voice and all seven durations across all six voices.

As for melodic complexity, the three levels of difficulty are determined by qualitative change. In the simple level only harmonic tones are used. In this way harmonic progression and melody are consistent and they make no deviations between synchronic and diachronic dimension. This helps to fuse together the distractor voices and makes it easier to perceive the target melody. In the medium level the distractor voices are composed using harmonic tones and non-harmonic tones on weak beats (e. g. passing notes). In this level each line assumes a stronger melodic character, becoming more salient in its own right, and drawing more attention from the listener. Finally, in the complex level, we find harmonic tones, non-harmonic tones on weak beats and non-harmonic tones on strong beats (e.g. appoggiaturas). In this last case, in contrast with the simplest level, harmonic progression and melody are not always consistent. This provokes some conflicts between harmony and melody, potentially drawing the attention of the listener away from the target melody in the strongest way possible.

Finally, for harmonic progression the difference between the levels is both qualitative and quantitative. In each level and each stimulus, seven different harmonic species are used. They are: major triad (I, major third + minor third), minor triad (VI, minor third + major third), diminished triad (VII, minor third + minor third), first species seventh chord (V7, major third + minor third + minor third), second species seventh chord (II7, minor third + major third + minor third), third species seventh chord (VII7, minor third + minor third + major third) and fourth species seventh chord (IV7, major third + minor third + major third). Each species is linked to a particular harmonic function: the I chord is linked to a tonic function; the VI chord is linked both to a subdominant and a tonic, depending on the context; VII, VII7 and V7 chords are associated with dominant function; while II7 and IV7 are exclusively associated with a subdominant function. In accordance with these characteristics, the three levels of difficulty are created using an increasing number of chords (quantitative difference), which implies an increasing number of different harmonic functions (qualitative difference), from the simplest level to the most complex one. The complexity of harmonic progression is hence created through an intensification of the harmonic functions and harmonic rhythm.

Overlap between voices was avoided as much as possible and each voice was composed in vocal style to enhance the independence of each melodic line. Where overlap was not avoidable it was usually sufficiently distant in terms of time (i.e. more than one measure away) not to disturb the independence of each voice.

As the three parameters can vary independently on three levels of difficulty each, 27 different stimuli were created, using Apple Logic X Software with the built-in sound library supplied with the program. The balance between the channels was regulated in order to have the target melody, played by a clarinet, and the two central voices, medium-high, played by a flute, and medium-low played by a cello, equally distributed on left and right channels. The very-high voice, played by a flute, and the very-low, played by a bass organ, were louder in the left channel; while the high, played by a flute, and the low, played by a cello, were louder in the right one. This was intended to spatialize the high and low register, while keeping the medium range equally balanced in the center. Each voice was manually adjusted by the first author to be of perceptually equal loudness. A -10 dB linear fade-in was applied to the six distractor voices on the first bar only to prime participant focus to the target melody (also determined through pilot testing). Stimuli were randomly produced in one of three slightly different tempos (87, 90 and 93 BPM) and three different keys (Bb Major, C Major, D Major) to avoid training effects as much as possible. All stimuli can be found online at *http://music-cognition.eecs.qmul.ac.uk/data/BussiSauvePearce2016.zip*.

C. Implementation of corpus analysis

1) Onset and offset synchrony. Onset synchrony (Duane, 2013) was calculated as the percent of pitches in a bar that have the same onset. In our implementation, the proportion of streams (out of the six distractor streams) with the same onset was averaged across every possible onset, in other words every sixteenth note (256 units of MIDI time), beginning with time 0 to produce one value for the composition. Offset (i.e. note ending) synchrony was calculated in the same way, beginning at time 256 (i.e. the end of the first possible 16th note).

2) *Pitch comodulation.* Pitch comodulation (Duane, 2013) was calculated for every vertical pair of pitches as it moves to another pair, with the following rules: a) when the two neighbouring harmonic intervals are the same, the scoring function should equal 1; b) the scoring function should decrease as the difference between the two intervals increases since larger interval differences are progressively further from parallel motion. The scoring function is as follows:

$$\text{score} = 1 - e^{\alpha/d} \quad (1)$$

where d is the absolute difference in semitones between the two harmonic intervals and α is a scaling constant, set to 2ln0.5 so that interval differences of 0 scored at 1, differences of 1 scored as .75, differences of 2 at .5 and so on. In our implementation, pitch comodulation was calculated for every possible pair of lines amongst the six distractor streams and averaged to produce a score for the composition.

3) *Harmonic overlap.* Harmonic overlap (Duane, 2013) calculates the degree of overlapping harmonics between two pitches (interval score) and is weighted by rhythmic duration. The interval score is calculated as follows:

$$\text{IntervalScore} = \frac{2}{Numerator + Denominator} \quad (2)$$

where the numerator and denominator are those of the interval's just frequency ratio (i.e. octave = 2/1, fifth = 3/2). An interval score is calculated for each pair of pitches with identical onsets between any two given streams. In our implementation, harmonic overlap is calculated in the same way as pitch comodulation: interval score is multiplied rhythmic weight, where a quarter note is equal to 1, a half note to 2, an eighth note to .5 and so on, for every possible pair of lines amongst the six distractor streams and averaged to produce a score for the composition.

D. Procedure

After reading the information sheet and providing informed consent, each participant was acquainted with the task and participated in a practice session, with an opportunity to ask questions before beginning the experiment. Participants listened through headphones (Bose, Sound True In-Ear) at a comfortable volume and sat at a comfortable distance from a laptop computer. The task was a same/difference judgment where participants were to indicate whether the target melody they heard presented in a polyphonic context was the same as the target melody they heard in a monophonic context. The monophonic melody was presented at the very beginning of the experiment and was present throughout so that participants could remind themselves of what it sounded like when they wished, but did not need to hear it for each trial (decided through piloting). In total, participants were presented with 2 excerpts of each of the 27 stimulus types for a total of 54 trials. One of each pair of excerpts contained the same target melody, while the other was altered in one of the four possible ways explained above. Moreover, in order to have an equal number of trials for each type of target melody change across the whole experiment, participants were randomly divided into 4 groups (8 subjects per group, 4 musicians and 4 non-musicians), in which the four alteration types were rotated. The 54 stimuli were randomized and the subjects could hear each stimulus a maximum of twice. The experiment was implemented in Max MSP and was run on an Apple MacBook Pro laptop computer.

III. RESULTS

A. Behavioural experiment

Accuracy was measured as a correct identification of *same* or *different* for each trial. Results showed a relatively low overall accuracy in completing the task, with a mean of .66, 95% CI [.64, .69], where musicians, with mean .76, [.73, .79] differed from non-musicians, with mean .56, [.53, .60]. A mixed-effects logistic regression predicting participant accuracy with type of target melody alteration as a 5-factor fixed effect (4 types plus no change), musicianship as a fixed effect, and participants as an intercept-only random effect yielded the best model, $\chi2$ (5) = 51.12, $p < .01$ as compared to a null model with only random effect of participant as predictor. See Table 1 for details of the model. A potential interaction between type of target melody alteration and musicianship was explored and while the model was significantly better than without an interaction, $\chi2$ (4) = 13.54, p = .008, interaction coefficients were highly correlated with corresponding type of target melody alteration coefficients and were therefore dropped. We can see by looking at the coefficients in Table 1 that types of melody change 2 and 4 contribute significantly to the model as compared to typologies 1 and 3, corresponding to changes in the last measure of the target melody as opposed to the middle. Musicianship is also a highly significant predictor, indicating that musicians and non-musicians performed differently, as outlined in overall performance above.

Table 1. Model specifications for behavioural and computational analysis models.

Model	Fixed effect	Estimate	Standard Error
Behavioural	(Intercept)	.12	.13
	Type1	.03	.16
	Type2	.48	.17
	Type3	-.05	.16
	Type4	.87	.18
	Musicianship	.99	.18
Computational	(Intercept)	6.20	2.62
	Harmonic Overlap	-3.83	1.69
	Musicianship	.98	.18

B. Computational analysis

In a similar analysis to Duane (2013), onset synchrony, offset synchrony, harmonic overlap and pitch comodulation were calculated as outlined above and tested as fixed effects to predict participant accuracy, along with a fixed effect of musicianship, and an intercept-only random effect of participants. The best model, as compared to a null model with only random effects of participants, contained harmonic overlap and musicianship as predictors as well as an intercept, $\chi 2$ (5) = 5.06, p = .02. See Table 1 for details of the model. Once again, the interaction between harmonic overlap and musicianship was significant, but the interaction coefficient was highly correlated to harmonic overlap and was therefore dropped.

IV. DISCUSSION

Overall, task accuracy was lower than targeted (~75%), particularly for non-musicians, who performed only slightly, though significantly above chance. This outcome is likely due to the difficulty of detection of the small differences in the target melody required by the task for non-musicians, and does not allow us to infer any explanation on how melodic, harmonic and rhythmic information is processed in this context.

Despite this limitation, the current study still provides some interesting results and implications. First it corroborates the importance of musical training in melodic segregation, as previously suggested (Marozeau et al., 2013; Zendel & Alain, 2009). Second, we have, to our knowledge, for the first time assessed stream segregation with complex and ecological polyphonic musical stimuli that systematically vary complexity in three musical parameters. Our results are contrary to predictions by Prince and colleagues (2009), which would suggest that pitch is a more salient feature than time because of its inherently greater complexity. This may be due to deliberate orthogonal manipulation of pitch structure and rhythmic structure in our stimuli. The results are also contrary to those of Duane (2013), who found evidence for the importance of onset synchrony and offset synchrony and pitch comodulation and none for harmonic overlap in the stream segregation of music. We suggest that this is due to the greater polyphonic complexity of our stimuli, both in terms of number of voices and harmonic complexity. However, it seems likely that the greater complexity of our stimuli also had a negative effect on task performance suggesting the need for larger alterations, greater separation between the voices or fewer voices to make the task easier in future research. Third, the location of the alteration had a significant effect on detection, where changes at the end of the target melody were more easily detected than those in the middle. This is likely to be a direct consequence of musical phrasing, where phrase endings are typically familiar, formulaic closures culminating on more or less definite cadences recognized across cultures (Huron, 1996), simplify detection of differences. This possibility is reinforced by both rhythm- and pitch-altered endings violating rules of musical closure.

ACKNOWLEDGMENT

This work is funded by Erasmus+/KAI1 Istruzione Superiore Mobility Consortia - North South Traineeship (2014/53), awarded to author MB.

REFERENCES

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press.

Duane, B. (2013). Auditory Streaming Cues in Eighteenth- and Early Nineteenth-Century String Quartets: A Corpus-Based Study. *Music Perception: An Interdisciplinary Journal*, *31*(1), 46–58. http://doi.org/10.1525/mp.2013.31.1.46

Huron, D. (1996). The melodic arch in Western folksongs. *Computing in Musicology*, *10*, 3–23.

Iverson, P. (1995). Auditory stream segregation by musical timbre: effects of static and dynamic acoustic attributes. *Journal of Experimental Psychology. Human Perception and Performance*, *21*(4), 751–763.

Marozeau, J., Innes-Brown, H., & Blamey, P. J. (2013). The Effect of Timbre and Loudness on Melody Segregation. *Music Perception: An Interdisciplinary Journal*, *30*(3), 259–274. http://doi.org/10.1525/mp.2012.30.3.259

McDermott, J. H., Lehr, A. J., & Oxenham, A. J. (2008). Is Relative Pitch Specific to Pitch? *Psychological Science*, *19*(12), 1263–1271. http://doi.org/10.1111/j.1467-9280.2008.02235.x

Prince, J. B., Thompson, W. F., & Schmuckler, M. A. (2009). Pitch and time, tonality and meter: how do musical dimensions combine? *Journal of Experimental Psychology. Human Perception and Performance*, *35*(5), 1598–1617. http://doi.org/10.1037/a0016456

Schoenberg, A., & Stein, L. (1969). *Structural functions of harmony*. WW Norton & Company.

Van Noorden, L. (1975). *Temporal Coherence in the Perception of Tone Sequences*. Technical University Eindhoven, Eindhoven.

Vliegen, J., Moore, B. C. J., & Oxenham, A. J. (1999). The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task. *Journal of the Acoustical Society of America*, *106*(2), 938–945. http://doi.org/10.1121/1.427140

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt, II [Investigations in Gestalt Theory: II. Laws of organization in perceptual forms]. *Psychologische Forschung*, *4*, 301–350.

Zendel, B. R., & Alain, C. (2009). Concurrent sound segregation is enhanced in musicians. *Journal of Cognitive Neuroscience*, *21*(8), 1488–1498. http://doi.org/10.1162/jocn.2009.21140

Reading Pitches at the Piano: Disentangling the Difficulties

Marion Wood

Institut für Psychologie, Westfälische Wilhelms-Universität, Münster, Germany
mwood@uni-muenster.de

ABSTRACT

To disentangle the cognitive processes involved in reading the pitch component of piano music, this research uses simple Reaction Time (RT) paradigms, and draws on task-switching and visual processing research. Participants were 44 pianists of widely varying age and competence. Pitches were shown three at a time and executed as fast as possible at a MIDI keyboard, with the key signature changed after 40-60 trials. A variety of specific factors could be quantified and contrasted, ranging from the cognitive 'switch-cost' of switching between clefs, to the extra processing time required when notes are presented in a different order. Large effects of clef and key signature persist in even the most accomplished professionals, adding 20% to response times at the very least, despite thousands of hours of practice. This paper tackles some of the questions about why this might be so. The extent to which overlapping, multivalent notation may obstruct both immediate execution and the longer-term learning process is discussed.

I. INTRODUCTION

Watching a pianist playing in a recital, often without any printed music, an audience member is often struck by the physical characteristics of the performance – the ability of the pianist to locate and play many notes in a complex and apparently effortless sequence. At the other end of the professional spectrum is the 'reluctant organist' attempting to play four hymns in a church service, with or without the added difficulty of coordinating notes on a pedalboard. We could be forgiven for thinking that motor coordination is the primary issue in studying a keyboard instrument. However a major issue preventing the exploration of a wide range of music that lies *within* the musician's technical grasp appears to be the difficulty of reading the printed notation. See Fourie, 2004, for example.

One of the problems in studying music reading is the interaction of the temporal (rhythmic) and physical (pitch) dimensions. This study takes as its focus the pitch aspect alone. By omitting the rhythm dimension, it is possible to apply insights from the considerable literature of existing response time (RT) research to the issue of music notation.

Musical notation, considered as a semiotic system, is not a very effective map of the physical space of the piano keyboard. It does not illustrate the octave-repeating pattern of the keyboard, and identical visual symbols or clusters of symbols must be executed differently by the two different clefs/hands.

Existing research into piano sight-reading (Sloboda, 1974) suggests that expert sight-readers may process common musical configurations as 'chunks' to a greater extent than novices. However as we see below, the notation system is not contributing effectively to the recognition of common musical patterns.

Tabulating the different possible responses to a single common triad, for example, (Figure 1) there are no less than ten visual-to-spatial mappings, considered across two clefs and eleven key signatures. The mappings also have different musical meanings: major, minor and diminished are words describing the musical 'character' of a chord. They all sound different, despite looking the same. The notation is not reflecting differences in execution mapping, auditory mapping, or musical meaning.

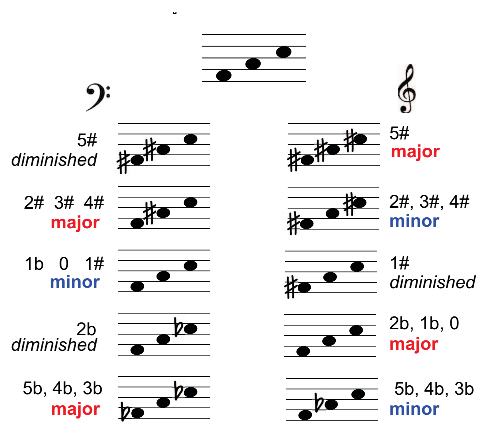


Figure 1. Ten different musical 'meanings', each with a specific motor response pattern, represented by a single visual fragment.

Research into the mental chronometry of visual processing suggests that 'multivalent' stimuli – *i.e.* visual symbols that can have more than one possible meaning – are processed more slowly than univalent symbols. Frequent changes to the response rules ('task-switches') slow the response time further, particularly in the first response after a switch. (Vandierendonck 2010).

Eye-tracking work by Goolsby (1994) showed pianists' gaze moving between treble and bass clef very frequently, sometimes on alternate saccades. If the two rule sets for treble and bass clefs can be considered a task-switch, clearly this would have implications for the speed of responding in piano sight-reading.

II. METHOD

Three separate experiments used a similar experimental setup, with three pitches presented to a computer screen at a time. The participant was requested to respond by playing the notes in the correct order, but as quickly as possible, as soon as they saw the stimulus. The key signature was given at the beginning of a block of 40-50 trials, and feedback on average response time and errors was given at the end of each block.

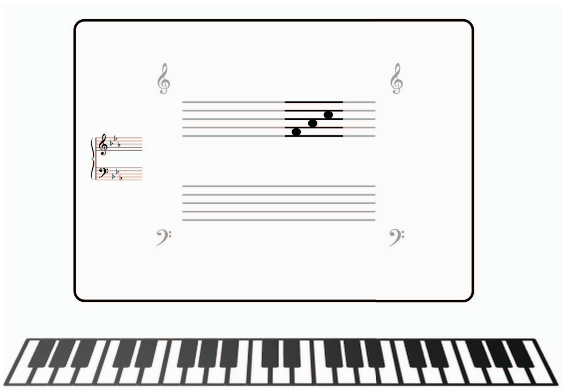


Figure 2. Experimental setup with screenshot of a repeat trial in the treble clef. A reminder key signature remains at the left during the whole block.

A. First Experiment

The first experiment used a classic alternating runs task-switching paradigm (Monsell 2003), in which two trials were presented to one hand/clef, followed by two trials to the other. Task-switch trials (immediately following a change of clef) may then be compared with task-repeat trials to measure the effect of the changing the task. A set of visually balanced triads was used as stimuli, shown in figure 3. This total set was subdivided so that not every triad appeared in every block, and the key signature was changed every two blocks. Thus a contrast could also be made between triads that had been recently rehearsed in the same key, or had recently appeared in a different key.

The whole experiment contained 18 blocks covering 9 keys, and were presented in two main key signature orders. 22 participants of widely varying age and musical background took part, recruited by word of mouth from in and around Exeter, UK, and their data were partitioned later into two equal-sized groups, of expert or moderate proficiency.

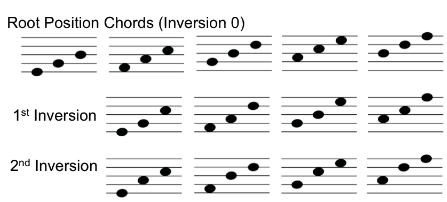


Figure 3. The stimulus set of experiment 1.

B. Second Experiment

The second experiment removed the task-switching element of the design (which although significant, was not found to be large in comparison to other effects) and also presented a more rigorously balanced set of experimental stimuli (figure 4), using seven root position chords. Every stimulus appeared in every block of every key. In this experiment each triad was also presented in reverse order. Nine key signatures were given in a randomly shuffled order.

Participants were all attending Dartington Summer School and were recruited through the summer school choir, so that in addition to their piano skills they were also all reasonably experienced choral singers. 12 participants had average RTs that approximately matched those of the moderate group from the previous experiment, and their data was analysed for comparison.

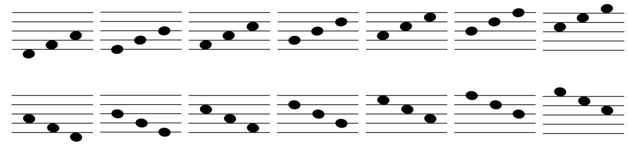


Figure 4. The balanced stimulus set of experiment 2.

C. Third Experiment

A third experiment focuses on the more difficult key signatures of 2#, 3#, 4# and 2b, 3b, 4b. Taking the same basic stimulus set as the second experiment, alternate blocks in the third experiment use a form of clarified notation (figure 5) that resolves the confusion between multivalent stimuli. 15 blocks present the six difficult key signatures (plus the central key C major/A minor) in one of four maximally confusing key orders, to try to evaluate the effect of one key signature on another, and establish to what extent visual ambiguity adds to response time.

Data from an initial group of 10 moderate participants is reported here, recruited from in and around Münster, Germany.

Figure 5. Modified noteheads, showing the chords Bb major and D major. Notes with the left half filled are flat (b), and with the right half filled, sharp (#). Short barlines clarify the clefs.

III. RESULTS

A more detailed account of these results may be found in the proceedings of the TENOR 2016 conference, where the the effects and applications of clarified notation are further discussed. Results are summarised here in brief, framing a discussion of the confounds in standard music-reading.

The data analysis relies on averaging the mean RT over groups of participants across the cells of the design. Although it would be possible to normalise the data across all participants, there are some aspects of motor coordination and cognitive architecture which are common to all levels of competence. Reaction time is a direct reflection of a physical quantity (processing duration) and is consequently not usually transformed in reporting experiments of this type.

Response time was taken at the third keypress, and averaged data across correct trials for analysis.

Error scores of -1 were mostly single errors of execution in the correct hand in the right general area of the keyboard, whereas errors of type -3 were almost all mistakes of switching (the wrong hand used, or wrong clef read).

A. Average Response Times

Participants in the ‘expert’ group of first experiment had average response times between 800 and 1500ms, and those in the ‘moderate’ group had response times between 1500 and 2500ms. Participants from the second experiment classed as ‘moderate’ had response times between 1150 and 2300ms, and from the third experiment between 1400 and 2450ms.

B. Effect of Clef

The effect of clef was highly significant across all experimental and pilot data, with the Treble clef generally faster than the bass clef. This Treble clef advantage included a number of self-reported left-handers, bass singers and 'cellists who might be expected to read bass clef more fluently.

In the expert group of the first experiment the contrast between the averages for each clef was 118ms, and in the moderate group it was 209ms. In the moderate group of the second experiment the impact of clef was proportionally lower but still highly significant, at 105ms. The contrast in the third experiment was 202ms, closely comparable to the moderate group of experiment one.

C. Effect of Switch of Clef

During the first experiment, in the expert group in particular, a number of very experienced individuals found it very hard not to alternate between the clefs/hands, and frequently hesitated or moved the wrong hand slightly in repeat trials.

Nevertheless a small but highly significant effect of 'task-switch' was found, in which the average response time when the clef had just been changed was greater than the average response time in trials where the clef remained the same. This difference was 46ms in the expert group, and 138ms in the moderate group.

D. Effect of Last-seen Clef Congruence

In the first experiment, the effect of stimuli on one another within the experiment was analysed in two ways.

1) On a global scale, three subsets of stimuli were rotated so that half the trials in each block were from a 'repeat set' – i.e. they were also shown in the previous block, and half from a 'novel set' that had been absent in the previous block.

2) At the local level, within each block, each stimulus appeared four times, once in each clef-switch/repeat condition, i.e. twice in each clef in each block. Investigating whether the RT of a stimulus is affected by its most recent previous appearance, trials were coded according to whether the stimulus had most recently been seen in the same (similar) clef, or in the other clef (different): see Figure 6.

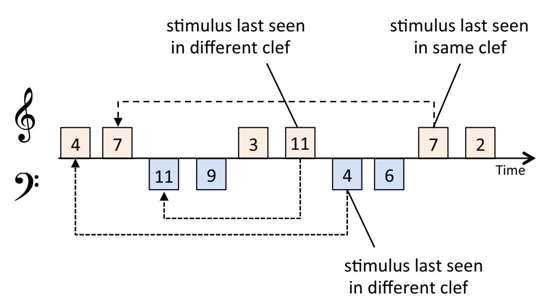


Figure 6. Illustration of last-seen-clef similarity. (Stimuli numbered arbitrarily).

Comparing the two subsets of 'novel' and 'repeat' stimuli within blocks where the key signature remained the same, no significant effect was found in either the expert or moderate groups, or in the error rates. The variable describing 'last-seen-clef' congruence, however, was found to be highly significant in both expert and moderate groups, with contrasts of 45ms and 96ms respectively.

E. Effect of Presentation Order

In the second and third experiments, the order of presentation was varied, with average response times to rising triads found to be faster than falling triads. In the second experiment this difference was 100ms, and in the third, 145ms.

F. Effect of Inversion

In the first experiment the effect of inversion was unexpectedly found to be significant, with no interaction with the number of black notes, or their position in the triad. In the expert group, this effect was small but highly significant, with a difference between the root and the first inversion 8ms, and between the first and second inversions 28ms, a combined difference of 29ms. In the moderate group, this effect was much larger, with a difference between root and first inversion of 168ms and between first and second inversions of 48ms, or 216ms combined.

Closer inspection of individual keypress data suggests that this may relate to the experimental procedure; the participants tended to arpeggiate their response in a single 'gesture'. The greater distance between two wider-spaced notes of an inverted chord is reflected in a slight delay between the keypresses. This effect appears to account for about half of the variation in inversion response times.

G. Effect of Diatonic Chord

The position of the chord within the key can be analysed according to its diatonic function, with chord 1 (I) the major key chord, and chord 6 (vi) the minor key chord. Chords 1-6 all appear in different positions in neighbouring keys, but chord 7 (*vii*), the diminished chord, is unique to each key signature and thus less practiced across the literature and within this experimental paradigm.

The contrast between chord 7 and the rest was significant in every group. The difference between the average response time for chord 1 and the average response time for chord 7 in the first experiment was 143ms in the expert group and 182ms in the moderate group; a proportionally much greater difference for the experts. In the second and third fully-balanced experiments this range was 219ms and 232ms respectively, with some pairwise comparisons between chords also significant.

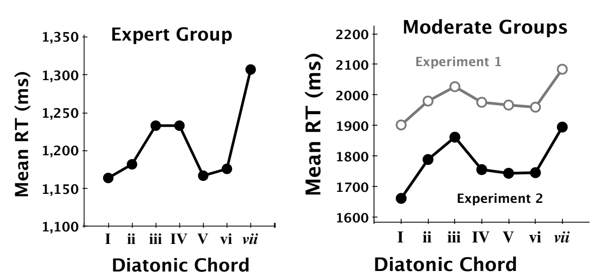


Figure 7. Diatonic Chord profiles

H. Effect of Key Signature

The contrast in performance between different key signatures was by far the largest effect seen in these experiments. There was a great variation in individual key signature profiles, but even in the expert group of the first experiment no participant showed less than 22% total variation across all nine keys, with others closer to 50%.

The first experiment was unbalanced by the unexpected effect of inversion, and a substantial effect of diatonic chord, neither of which had been allowed for in the selection of stimulus sets. In addition, the experts seemed to have continued to improve their performance beyond the practice blocks and into the first block of the experiment, which was either 2# or 2b. Consequently their key signature profiles, whilst generally indicative, may not be entirely typical. The second experiment was more rigorously balanced and shows a an approximately symmetrical shape, with key signatures gradually increasing in mean response time as the number of sharps or flats increases. Three modifiers seems to be an exception, with 3b showing what appears to be an advantage, and 3# a disadvantage. The third experiment, still in progress, sets out to investigate this apparent discrepancy further.

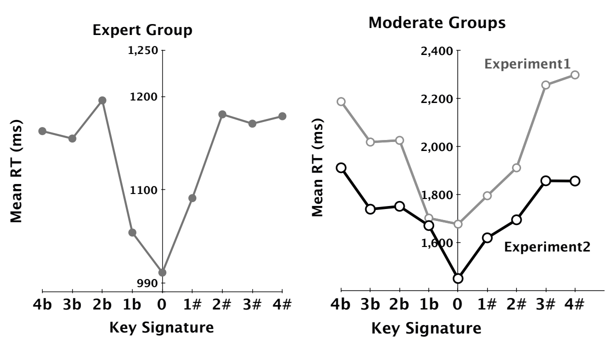


Figure 8. Key signature Profiles

I. Effect of Accidentals

An analysis of the number of black notes in each trial was used to test whether this factor could by itself account for the variation in key signature data. In the first experiment, in the expert group only, there was a moderately significant delay in executing chords with two black notes, but this effect was small in comparison to the general overall effect of keysignature. In the later experiments, although the same contrast showed signs of approaching significance, the effect was again very small, and in no sense accounts for the wide variation across key signatures.

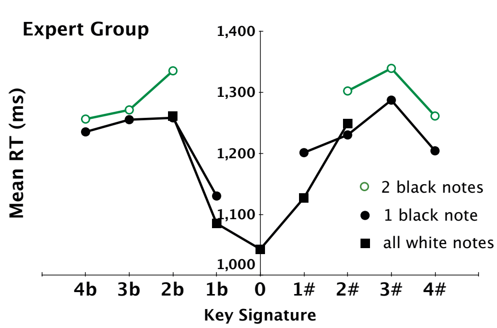


Figure 9. Mean RT in the Expert group of chords with 0, 1 and 2 accidentals.

J. Settling-in Effect

In the first experiment, in which two blocks of each key signature followed one another, the second block of each key was found to be faster than the first block. A finer-grained analysis found this settling-in effect took place mostly in the first third of a block: *i.e.* the first 11 or 12 trials.

K. Effect of Clarified Notation

The third experiment investigates the more difficult keys of 2, 3 and 4 modifiers and their effect on one another in close proximity. To elicit more information about the role of visual confusion in the notation, alternate blocks are given in the clarified notation of figure 3.

The effect of this notation was itself highly significant, with the average response time in clarified blocks 277ms faster than in the traditional blocks. In addition, a dramatic fall in the error rates was observed, particularly in 1-note errors. For more discussion, see the proceedings of the TENOR 2016 conference (in press).

For the current discussion, although the data presented here constitutes only one quarter of a larger 4x4 design, initial analysis suggests that 3# remains more difficult than 3b, and that neither the learning effect provided by an intensive balanced design, or the effect of clarified notation seems to have entirely dispelled this tendency. On the other hand, clarified notation seems to have mitigated about half the total effect of key signature.

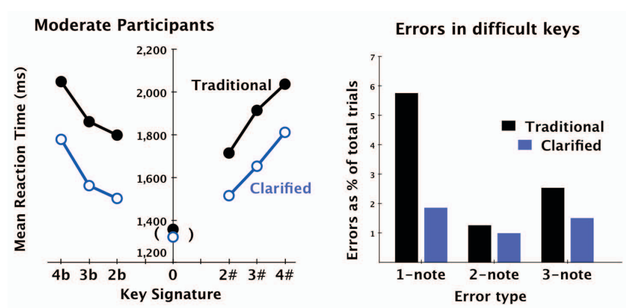


Figure 10, Effect of Clarified Notation

L. Summary of Results

Table 1. Summary of effects across three experiments. All times given in ms.

	Experiment 1		Exp. 2	Exp. 3
Effect	**Expert**	**Mod.**	**Mod.**	**Mod.**
Range of average RTs	*800-1500*	*1500-2500*	*1150-2300*	*1400-2450*
Clef	118	209	105	202
Switch of clef	46	138		
Key signature	185	620*	460	575
Effect of key change	65	111*		
Last-seen-clef	45	96		
2 black notes	41	(n.s.)*		
Inversion	29	216		
Diatonic Chord	143	182*	219	232
Key Signature	185	620*	460	575
Note Order			100	145
Notation				277

* Effects that may be unbalanced by the effect of inversion.

IV. DISCUSSION

In principle there may be a number of factors at work to explain the variation in response times reported here. Hand-eye coordination in this situation requires at least three steps: decoding of the instructions, some transformation into an action plan, and then the execution of that plan in space. Delays may be incurred by extra difficulties at any of these stages.

The only evidence for significant motor difficulty in executing common chords at the keyboard was provided by the very small delay for chords with two black notes in the expert group of the first experiment. The contrast between clef/hand is unlikely to be a coordination problem because it remained evident in left-handers.

A likely example of a transformation difficulty is the effect of presentation order (III.E). In the context of piano playing, ascending figures are congruent to the order in which the notes are to be executed, shown in figure 11 below. Processing notes from left to right and then executing them in the other direction requires a mirror transformation that may confer a delay.

The difference between the clefs themselves can also be considered from this point of view. The two hands are mirror images of one another, operating in an environment that repeats consistently from left to right. So either a chord or a visual pattern that is fingered 1-3-5 in the right hand, will be executed 5-3-1 in the left hand; a similar mirror transformation. This would not explain why the right hand/treble clef is faster in left-handed individuals, but provides one hypothesis for the difference between hands.

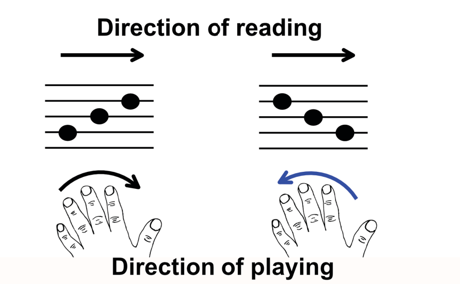


Figure 11. Playing/Reading incongruence in falling triads

In terms of decoding the instructions, several factors may be at work. We expect a two-fold effect of multivalent notation, both in obscuring the actual requirements at the point of execution, and in obstructing the long-term pattern learning that we know is a pre-requisite for good sight-reading. Disentangling these effects from one another, and from middle-term recency effects of one key signature on another is a complex but fascinating challenge. Pianists represent a population with thousands of hours of practice at a very complex and interlocking set of task-mappings, which is unmatched in normal experimental settings.

The effect of last-seen congruence reported in section III.D is an important finding. Apparently the clef of the last sighting of the a visual configuration (either in the same clef or the other clef) was more significant than whether that stimulus had been rehearsed in the previous block, where it would have been shown four times, twice in each clef. This suggests that beyond the small 'task-switch' effect, the switch might in itself have the capacity to disrupt recent learning.

In terms of longer term learning, there are two main lenses we might apply. The 'cellists (including one who was also left-handed), on being informed that they read more fluently in the right hand expressed little surprise, stating that they had learned the Treble clef first as children, and had always felt comfortable reading it. If early experience of a set of task rules forms the basis for music-reading, with other mappings either overlaid in conflict, or extended by some further transformation, we might also be able to explain the overwhelming advantage observed for the central key of C major/A minor, which most piano teachers introduce first.

On the other hand, accumulated exposure may play a substantial role in shaping long term learning. The treble clef effect may simply reflect the fact that the timbral qualities of the instrument mean there are usually more notes to play in the right hand than the left, those notes are somewhat less likely to form patterns that can be taken in at a glance, and consequently treble clef reading is more practiced. Whether the same can be said of the central key C major/A minor is more questionable, as any young learner who has spent many hours playing the Moonlight Sonata will attest.

There is some indication in the data that some keys may be read as 'extensions' of other related keys. Some 50% of all the single-note errors in the first and second experiments were semitone errors at the 4^{th} or 7^{th} degree of the scale, i.e. forgetting the last flat or sharp of the key signature. This seemed to be irrespective of the immediately preceding key signature. In addition, some of the diatonic graphs showed signs of skew towards a particular reading strategy. When chord IV appears to be the most fluently read in the key of 1#, for example, we may suppose that the key is essentially being read as "Cmajor with an extra sharp". Additionally some participants' individual graphs in 3b show an advantage for chords vi and iii (when 3 is usually the most disadvantaged chord after chord 7) which strongly suggests that 3b is being read preferentially as C minor, rather than as Eb major. These 'extension' strategies appear to be quite individual, however, and may reduce or disappear in experts; more data from balanced experiments is required to investigate this effect.

Key signatures may interfere with each other in a number of ways. One might imagine that A major (3#) followed by A flat major (4b) might present a particular set of difficulties, as every single note mapping will be altered. On the other hand, the diatonic chords remain in the same locations on the musical staff, which may be an advantage. In terms of pure spatial congruence, 3b is the mirror inverse of 3#, and playing these one after the other might also be expected to cause a particular pattern of disruption between the hands. A further possibility, that the major and minor key signatures of the same keynote either enable or disrupt one another could also be considered. None of these questions provide a direct hypothesis for why 3# should show longer average response times than 3b, but provide plausible avenues of enquiry which form the basis of the third experiment.

V. CONCLUSION

The multivalent nature of music notation, as discussed in the introduction, may provide a reasonable hypothesis for the particular difficulty of sight-reading in two clefs and in many keys. This experiment was set up to encourage 'chunking' of three-group notes into patterns that are amongst the most common in the literature, and to observe how better pattern recognition of these chunks may be acquired.

Further work aims to collect more data from expert participants to contrast with the moderate group of the second experiment, and to complete the key signature comparison design of the third experiment. Further study of the ways in which individuals select or extend core key signature mappings, and of the learning that takes place during the experiment is also planned.

ACKNOWLEDGMENTS

Many thanks to all the participants for their time, interest, suggestions and anecdotal insights, to Professor Stephen Monsell for his supervision in devising the experiment, and to the staff of Dartington Summer School for making the second experiment possible.

REFERENCES

Fourie, E. (2004). The processing of music notation: some implications for piano sight-reading. *Journal of the Musical Arts in Africa.*, *1*, 1–23.

Goolsby Thomas W. (1994). Eye Movement in Music Reading: Effects of Reading Ability, Notational Complexity, and Encounters. *Music Perception: An Interdisciplinary Journal*, *12*(1), 77–96.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, *7*(3), 134–140. http://doi.org/10.1016/S1364-6613(03)00028-7

Sloboda, J. (1974). The Eye-Hand Span-An Approach to the Study of Sight Reading. *Psychology of Music*, *2*(2), 4–10. http://doi.org/10.1177/030573567422001

Vandierendonck, A., Liefooghe, B., & Verbruggen, F. (2010). Task switching: Interplay of reconfiguration and interference control. *Psychological Bulletin*, *136*(4), 601–626. http://doi.org/10.1037/a0019791

Wood, M (2016). Visual Confusion in Piano Notation. In *Proceedings of the Second International Conference on Technologies for Music Notation and Representation* (In press)

A Machine-Learned Model of Perceived Musical Affect

Tess Dula,[*1] Paul Griesemer,[*2] Joshua D. Albrecht[#3]

[*] *Computer Science, Engineering, and Physics Department, The University of Mary Hardin-Baylor, Belton, TX 76513, USA*

[#] *Music Department, The University of Mary Hardin-Baylor, Belton, TX 76513, USA*

[1]tadula@mail.umhb.edu, [2]pgriesemer@umhb.edu, [3]jalbrecht@umhb.edu

ABSTRACT

Recent research has shown that affect perception in music can be reliably correlated with low-level acoustic parameters. In this study, we examine thirty excerpts from the Beethoven piano sonatas to generate a cognitive model of affect perception in this repertoire. Specifically, six low-level acoustic features have been derived from each excerpt using automatic feature extraction, including timbre, maximum amplitude, average amplitude, frequency of the dynamic signal, rate of change of the spectral centroid, and roughness. These acoustic features are used to predict perceived musical affect from 68 participants in this repertoire by using machine-learned neural networks. Nine affective dimensions are tested, resulting in nine models that are able to model the perception of affect based on low-level acoustic features. Specifically, between 7.1% and 41.5% of the variance in listener ratings are able to be accounted for by the resulting cognitive models.

I. INTRODUCTION

There are an incredible variety of factors that influence how a listener perceives emotion communicated through a musical signal, including prior listening habits and exposure, personal disposition and personality, level of attentiveness, attitudes about the music heard, current mood, autobiographical associations with the music heard, age and gender, and the environment in which the music is heard, to name just a few (see Juslin & Västfjall, 2008). In short, perceptions of emotions expressed in music are personal and situation-specific.

With so many variables affecting perception of musical emotion, it is remarkable that any researchers would find any patterns shared between listeners about perceptions of musical expression. Despite the challenges, however, researchers have consistently shown that not only can generalizations be made about emotions perceived in music, but that in the right circumstances a relatively large amount of the variance in listener perceptions of musical emotion can be accounted for through relatively low-level musical parameters.

For example, by examining only pitch, mode, melodic progression, rhythm, tempo, sound level, articulation, and timbre, Juslin and Lindström (2010) were able to effectively model perceived emotion in music. In another study, Eerola, Friberg, and Bresin (2013) systematically manipulated mode, tempo, articulation, timbre, and register, and through these variables were able to account for 77-89% of the variance in listener perceptions of happy, sad, peaceful, and angry ratings.

In most environments, however, listeners have access to sound waves only and not the score. In short, when listening to music, listeners' brains receive as input sound waves and from that signal alone decode emotional expression. Given the cognitive processes at work, researchers should be able to effectively model musical emotion based only on the audio signals.

Using this process, Schubert (2004) was able to account for between 33% and 73% of the variance in continuously collected listener ratings of affect in music by examining only the loudness, tempo, musical contour, texture, and spectral centroid of musical excerpts. In another study, Eerola, Lartillot, and Toiviainen (2009) used automatic feature extraction to predict perceived musical emotion in film music extracts. Using only 2 to 5 predictors, they were able to account for between 58 and 85% of the variance in listener response along happiness, sadness, tenderness, fear, and anger dimensions.

In light of these results, automatic low-level musical feature extraction of audio signals promises to offer fruitful raw material for predicting and modeling listener perceptions of musical affect. In other words, based only on the input of the musical features embedded in an audio signal listeners' minds appear to be able to consistently assign expressions of musical affect as output in ways that are inter-subjectively reliable.

One way to gain an understanding of the cognitive processes involved in the perception of musical emotion, then, would be to use machine learning to explore the relationships between low-level musical features automatically extracted from an audio excerpt and the affect ratings assigned to that excerpt. The relationships discovered by a machine-learned neural network may offer insights into the cognitive processes listeners use to ascribe affective expression to musical excerpts.

II. AIMS

The purpose of this study is to explore relationships between low-level musical features automatically extracted from a set of musical excerpts and affect ratings of those excerpts provided by listeners. The end result of this study is to produce a machine-learned neural network that will offer a cognitive model of the way that listeners perceive affective communication of musical signals.

III. METHODOLOGY

The perpetual data used for the supervised training of the machine-learned neural network model was collected in a prior study. Specifically, Albrecht (2016) collected listeners' ratings of affect in a subset of excerpts taken from the Beethoven piano sonata literature. These excerpts were chosen as representative of some of the most powerfully expressive moments in the Beethoven piano sonatas. Perceptual data was collected along nine affective dimensions empirically determined to be relevant for this repertoire (Albrecht, 2014).

For the purpose of the cognitive model, six acoustic features were automatically extracted from each of these excerpts. As a first step in building a neural network examining the relationship between these acoustic features and listener ratings of affect, regression models were constructed. These regression models can be taken as proxies of correlations between the surface musical features and listener perceptions of affective expression.

A. Stimuli

For this study, it was important to use excerpts that would be representative of strong affective expressions. Therefore, the Society for Music Theory listserv, the American Musicological Society listserv, and the Piano Street forum were polled for those excerpts that were deemed to be "the most emotionally expressive moments in the Beethoven piano sonata literature." As a result of this poll, thirty excerpts were provided, along with the descriptions of the excerpts provided in parentheses.

To control for excerpt length, excerpts of five seconds in length were chosen. If the excerpt provided by the respondent was longer than five seconds, the five seconds including the moment most closely matching the description was chosen. If there was no description of the passage, the first five seconds were used. More detail on recordings chosen can be found in Albrecht (2016). The excerpts used are listed in Table 1.

Table 1. The thirty excerpts used in this study as a result of polling expert musicians.

Excerpt	Excerpt
Op. 2, No. 1, II (2nd theme)	Op. 53, III (major 7th arpeg.)
Op. 2, No. 1, IV (mm. 20-50)	Op. 57, I (start of 2nd theme)
Op. 7, II (2nd theme)	Op. 81a, III (ending)
Op. 10, No. 3, I (no section)	Op. 81a, II (2nd theme)
Op. 10, No. 3, II (development)	Op. 90, II (canon)
Op. 13, No. 1, II (entire mvt.)	Op. 101, III (into to final mvt.)
Op. 26, I (resolutions in Var. III)	Op. 106, I (beginning)
Op. 26, IV (main theme)	Op. 106, II (beginning)
Op. 27, No. 2, I (trans. to A1)	Op. 106, III (mm. 60 or 145)
Op. 27, No. 2, III (opening)	Op. 109, I (retransition)
Op. 28, I (mm. 109-127)	Op. 109, III (Var. VI)
Op. 31, No. 2, I (recit. at close)	Op. 110, I (sad part)
Op. 31, No. 2, III (mm. 169-173)	Op. 110, III (mm. 21-23)
Op. 31, No. 3, I (ms. 99)	Op. 111, I (beginning)
Op. 53, I (2nd theme)	Op. 111, II mm. (116-121)

B. Affect terms

Affect terms in this study were taken from Albrecht (2014), who conducted a series of three experiments to empirically determine which affect labels would be appropriate for studying the Beethoven piano sonatas. From these studies, he obtained nine affective dimensions, used in this study: angry, anticipating, anxious, calm, dreamy, exciting, happy, in love, and sad.

C. Participants

This study involved 68 participants from the University of Mary Hardin-Baylor, consisting of freshmen and sophomore music majors. Among these participants there were 33 females, a mean age of 19.4 (SD = 1.5), a mean number of years of musical training of 7.1 (SD = 3.7), and a mean Ollen Musical Sophistication Index (Ollen, 2006) of 455.0 (SD = 243.9). More detail about the participants can be found in Albrecht (2016).

D. Procedure

Subjects participated in groups of up to 8 in a computer lab. After reading the experiment's instructions, participants listened to two trial excerpts to ensure they understood the instructions and interface. Participants listened to the excerpts on headphones adjusted to a comfortable listening level, and rated each excerpt's affective expression on a 7-point Likert scale.

E. Automatic feature extraction

For the purpose of constructing models of musical affect, six acoustic features were automatically extracted from each of the thirty five-second audio excerpts. These features were extracted using LabVIEW 2015, version 15.02f. For the purpose of continuously tracking certain audio features across the five second excerpt, excerpts were subdivided into fifty 100-millisecond windows. A fast Fourier transform was performed to determine the amplitudes of the frequencies present in each window. The algorithms for the extraction of each of the six audio features are described further below.

1. *Timbre.* The spectral centroid, f_c, of each excerpt was used as a measure of the timbre of the piece (see Schubert, Wolfe, and Tarnopolsky, 2004). The fast Fourier transform produces a power spectrum that relates amplitudes of a signal, A, to the corresponding frequency, f. Using this information, the spectral centroid was computed as the weighted average of frequencies by amplitude,

$$f_c = \frac{\sum A(f_i) f_i}{\sum A(f_i)}.$$

2. *Maximum Amplitude.* The measurement of dynamic range was extracted from each 100-millisecond splice. Amplitude was recorded at the peak of each note-produced sine wave. The maximum amplitude of each splice was collected. The largest maximum of all of the splices was then selected as the maximum amplitude of the excerpt.

3. *Average Amplitude.* Using the power spectrum produced by the fast Fourier transform, the amplitude of the two strongest signals were acquired for each 100-millisecond window. These amplitudes were then averaged for each splice throughout the piece to gain a measured mean volume.

4. *Roughness.* Vassilakis (2005) showed that auditory roughness corresponds to auditory consonance. In this study roughness was measured within each 100-millisecond splice using the following formula from Vassilakis:

$$R = X^{0.1} * 0.5 * Y^{3.11} * Z,$$

where:

$$X = A_{min} * A_{max}$$

$$Y = \frac{2 * A_{min}}{A_{min} + A_{max}}$$

$$Z = e^{-b_1 * s(f_{max} - f_{min})} - e^{-b_2 * s(f_{max} - f_{min})}.$$

f_{min} and f_{max} are the minimum and maximum amplitudes of the two strongest frequencies in the power spectrum of each 100-millisecond splice. A_{min} and A_{max} are the minimum and maximum amplitudes, respectively, of the same two frequencies. X corresponds to the dependence of roughness on intensity (amplitude) of accounted sine waves. Y denotes the fluctuation degree of amplitude. Z represents the relation of roughness to the fluctuation rate of amplitude.

5. Spectral Centroid Rate of Change. As defined above, the spectral centroid was measured in each 100-millisecond splice. To find the spectral centroid rate of change, the following equation was applied to each consecutive pair of splices:

$$\sum |f_{c_i} - f_{c_{i+1}}|.$$

Spectral centroid rate of change measures velocity of spectral centroid throughout the piece—a calculation of the speed at which sound quality fluctuated throughout the five-second excerpt. The summation of individual splice measurements provided a numeric measurement for spectral rate of change over the entirety of the five-second musical excerpt.

6. Oscillation Mode of the Dynamic Signal. The final predictor in this study is the velocity of the amplitude signal throughout the excerpt. Amplitude measurements for each 100-millisecond splice were recorded and plotted against time. An example of this plot is shown in Figure 1 for Op. 31, No. 3, I. The resulting wave was analyzed with a Fourier transform, and the dominant frequency was recorded. This frequency is the principal mode of oscillation of the dynamic signal. For the example shown, the frequency was found to be 0.26 Hz.

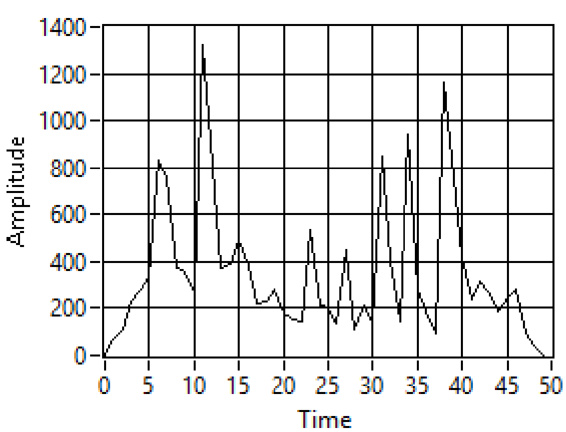


Figure 1 - Example of the Dynamic Signal for Excerpt Op. 31, No. 3, I, with a Frequency of 0.26 Hz.

IV. RESULTS

A. Predictor Selection

Once values were assigned to each excerpt according to the six low-level acoustic features defined above (see III: E), each variable was treated as a predictor variable in a linear regression model for each of the nine affective categories. Model selection is not a straightforward task, and suffers from dangers of overfitting and covariance. Regression models were selected that a) maximized variance accounted for, b) minimized overfitting by limiting the number of predictors, and c) selecting for theoretically meaningful predictors. One model was created for each affective dimension.

Derived models included 2-5 predictor variables each, accounting for between 6.9%-41.5% (mean of 20.9%) of the variance in participant ratings. The musical parameters used in each model, along with the directionality of the effect, are shown in Table 2.

Table 2. Regression models derived from the training set. Adj. R^2 values are what percent of the variance is explained.

Affective dimension	Regression model	Adj. R^2
Angry	Overall maximum amplitude, timbre, spectral centroid rate of change.	0.111
Anticipating	Overall maximum amplitude, oscillation mode of the dynamic signal, spectral centroid rate of change.	0.071
Anxious	Overall maximum amplitude, oscillation mode of the dynamic signal, spectral centroid rate of change.	0.082
Calm	Overall maximum amplitude, average amplitude, timbre, roughness, spectral centroid rate of change.	0.401
Dreamy	Overall maximum amplitude, average amplitude, timbre, spectral centroid rate of change.	0.287
Exciting	Overall maximum amplitude, average amplitude, roughness.	0.415
Happy	Overall maximum amplitude, average amplitude, roughness.	0.223
In love	Overall maximum amplitude, timbre, spectral centroid rate of change.	0.072
Sad	Average amplitude, roughness.	0.222

V. DISCUSSION

The results from this study are consistent with the hypothesis that low-level automatically extracted acoustic features can predict listener ratings of perceived affect in Beethoven's piano sonatas with a statistically significant accuracy. Nevertheless, the study's methodology and results should be contextualized with a few caveats.

A. Additional acoustic variables

In the first place, it is important to note that the six acoustic features examined in this study are somewhat limited. Several important elements of a musical signal are not fully accounted for by these predictors. For example, although average spectral centroid gives an approximation of timbre and pitch height, more detailed pitch information could be taken into consideration. The partials present in the signal could be algorithmically reduced to approximations of complex pitches made up of multiple partials in the cochlea. From this algorithm, a measure could be made of the highest and lowest pitches, as well as the density of chords.

Another variable that could be further explored is melodic progression. 100 ms. is a short time, but many musical passages use notes shorter than 100 ms. Further divisions in window size could be made allowing for a more detailed investigation of melodic contour and motion. A third variable measured could be mode, which has been shown to be a significant indicator of affective expression. Further research could quantify the 'majorness' or 'minorness' of each window length and so estimate the mode of each excerpt.

Finally, variables related to rhythm and tempo could be extracted from the signal. The number of onset moments in each five second excerpt could be calculated by looking at the inter-onset intervals between amplitude spikes, providing an estimate of rhythmic activity. In addition, a tempo variable could be algorithmically derived based on rhythmic and accentual regularities in the acoustic signal.

B. Model Strength

The 'missing variables' discussed in the preceding section almost certainly play a role in reducing the amount of variance in participant ratings the regression models reported here can explain. Of course, listeners who are parsing the acoustic signal in real time have access to all of the variables described in the preceding section, and the failure to include them in these models likely limits the success of the models' predictions.

Although each regression model listed can explain a significant amount of the variance in listener ratings, some of the models do a poor job of fitting the data. For example, the anticipating regression model can only account for 7.1% of the variance, while in love and anxious only account for 7.2% and 8.2%, respectively. It is likely that these numbers would increase with the inclusion of more variables.

Some models can explain up to 41.5% (exciting) or 40.1% (calm) of the listener variance. These higher values offer more confidence that the models are more accurately measuring musical parameters associated with these emotions. Nevertheless, in comparison to Eerola, Friberg, and Bresin (2013), Schubert (2004), or Eerola, Lartillot, and Toiviainen (2009), there is still room for improvement.

C. Overfitting

As mentioned above, the regression models derived from this dataset may be overfit to the data. Although adjusted R2 values attempt to correct for overfitting by penalizing extra variable selection, it is still possible that the models may be overly specific with respect to music participants, excerpt selection, and affective communication. Further research and algorithms will need to be developed in order to address these categories of overfitting.

VI. CONCLUSION

The purpose of this research effort is to construct machine-learning algorithms to analyze and predict the correlation between low-level musical features automatically extracted from an audio excerpt and participants' evaluation of the emotional expression of that excerpt. One way to begin an effort to build a machine-learned neural net exploring these relationships is to first investigate correlations using a multiple regression model. The regression models reported here show that the low-level acoustic features automatically extracted from these audio signals are significantly correlated with perceived emotion. It is therefore fitting to further expand this work by training a machine-learned neural network on the same task. It is possible that a machine-learned model may provide further insight into not only the relationships embedded in the signal, but also on the way humans process a cognitively complex signal like musical emotion.

In future work, the addition of further parameters will be explored due to the low adjusted R^2 values observed in the current set. Examples of further parameters include: melodic progression, rhythm, and tempo throughout the musical excerpt.

The next step in this research program is to define the machine learning algorithms that will employ the automatic extraction algorithms that are described in the current work. These algorithms will be trained on the data set described above and the results will be compared to current statistical approaches to the problem of modeling the emotional response to complex pieces of music.

REFERENCES

Albrecht, J. (2014). Naming the abstract: Building a repertoire-appropriate taxonomy of affective expression. In *Proceedings of ICMPC-APSCOM 2014 Joint Conference* (pp. 128-135). Seoul, South Korea: Yonsei University.

Albrecht, J. (2016). Learning the language of affect: A new model that predicts perceived affect across Beethoven's piano sonatas using music-theoretic parameters. in *Proceedings of the ICMPC-SMPC 2016 Joint Conference*, San Francisco, USA.

Eerola T., Friberg A., & Bresin R. (2013). Emotional expression in music: contribution, linearity, and additivity of primary musical cues. *Frontiers in Psychology*. **4**:487. doi: 10.3389/fpsyg.2013.00487

Eerola, T., Lartillot, O., & Toiviainen, P. (2009). Prediction of multidimensional ratings in music from audio using multivariate regression models. In *Proceedings of 10th International Conference on Music Information Retrieval (ISMIR 2009)* (pp. 621-626). Kobe, Japan.

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and brain sciences*, *31*(05), 559-575.

Juslin, P. N., & Lindstrom, E. (2010). Musical expression of emotions: Modelling listeners' judgements of composed and performed features. *Music Analysis*, *29*(1-3), 334-364.

Ollen, J. (2006). *A Criterion-related Validity Test of Selected Indicators of Musical Sophistication Using Expert Ratings* (Doctoral dissertation). Columbus, OH: Ohio State University.

Schubert, E. (2004). Modeling perceived emotion with continuous musical features. *Music Perception: An Interdisciplinary Journal*, *21*(4), 561-585.

Schubert, E., Wolfe, J., & Tarnopolsky, A. (2004, August). Spectral centroid and timbre in complex, multiple instrumental textures. In *Proc. 8th Int. Conf. on Music Perception & Cognition (ICMPC), Evanston*.

Vassilakis, P. N. (2005). Auditory roughness as means of musical expression. *Selected Reports in Ethnomusicology*, *12*, 119-144.

Patterns of Complexity in Improvisational Jazz Performance

Ashley Walton[1], Auriel Washburn[1], Peter Langland-Hassan[2], Anthony Chemero[1,2], and Michael J. Richardson[1]
[1]Department of Psychology, University of Cincinnati, Cincinnati, OH, USA
[2]Department of Philosophy, University of Cincinnati, Cincinnati, OH, USA

Interpersonal coordination is key for successful musical performance, demanding the collective synchronization of auditory and kinesthetic dimensions with respect to both the musical structure and the actions of co-performers. A fundamental aspect of such interaction is the bodily coordination that occurs among musicians. Fractal analyses provide a measure of complexity for such behavior, quantifying the self-similarity, or long-term correlations in movement time series. High degrees of self-similarity have been observed in a wide array of natural processes including heart rate variability and gait patterns, and are also exhibited in the structured evolution of inter-beat-interval timing and hand movements during musical performance. Additionally, previous work has established that the fractal structure, or complexity of movements is influenced by environmental constraints, and that there is often an association between the movement complexity exhibited by co-actors during interaction. This study examines how the complexity of musicians' movements changes as a function of the musical context, as well as how the complexity of their movements varies with that of their co-performers. Pairs of pianists improvised with different backing tracks (i.e., with high or low levels of harmonic and rhythmic structure), providing contrasting levels of constraint on performance. Results revealed more self-similarity in the musicians' left arm and head movements for the backing track with less structure, while movements produced during performance with the more structured backing track were characterized by less self-similarity. An assessment of the relationship between the self-similarity of the performer's movements was also examined in order to understand the need for complexity matching vs. complementarity during musical interaction. Lastly, listener evaluations will provide information about how reciprocal playing behavior relates to the aesthetic value of a performance.

The effect of intensive jazz improvisation instruction on measures of executive function in middle school band students

Martin Norgaard[1], Heston McCranie[1]
[1]School of Music, Georgia State University, Atlanta, Georgia, USA

Research shows that active music instruction with K-12 populations may enhance academic achievement. This enhancement may be due to better auditory processing in students who participate in music. We believe there are additional possible advantages to active music instruction with a focus on musical improvisation. Improvisation involves combining discrete elements (notes, musical figures) in real time following musical rules. In addition, improvisers plan architectural features of upcoming passages and evaluate whether the played output corresponds to these plans. The evaluation process also identifies errors in note choices that may not comply with the given tonal and rhythmic context. Therefore, we hypothesize that students who participate in intensive jazz improvisation instruction exhibit enhanced general cognitive abilities on measures related to inhibition, cognitive flexibility, and working memory. We are testing this hypothesis in an ongoing longitudinal study in which middle school students in a large suburban band program complete cognitive tests before and after receiving intensive jazz improvisation instruction. Initial between-subjects results show that students with improvisation experience (N=24) make significantly fewer errors on the Wisconsin Card Sort Task (M=11.9, % perseverative errors) than a control group (M=14.7, N=141). However, there were no significant differences between the two groups in measures of cognitive inhibition and working memory performance. We are currently engaging a much larger group in intensive jazz instruction (N=70). Results of these within-subjects pre/post tests will be available by the time of the proposed presentation. We believe this is the first investigation of far-transfer effects of music improvisation instruction.

Jazz Improvisation as a Model of the Creative Process: Heightened Perceptual Awareness and Sensitivity

Psyche Loui,* Emily Przysinda*, Pheeroan Aklaff#, Kellyn Maves*, Cameron Arkin*, Tima Zeng*

*Department of Psychology, Program in Neuroscience & Behavior, Wesleyan University, USA
#Department of Music, Wesleyan University, USA

ploui, eprzysinda, paklaff, kmaves, carkin, hzeng@wesleyan.edu

ABSTRACT

The process of creativity entails the production of novel and original work that takes into account the domain, the field, and the creator (Csikszentmihalyi, 1996). Here we report recent theoretical and empirical advances on jazz improvisation as a model for understanding the process of creativity. We propose a framework by which musicians can learn to become creative improvisers via simultaneous perceptual, cognitive, and social engagement. These learning processes translate to gaining active experience with musical structures (such as scales and chords), exposure to established works in the field, and ensemble improvisation with musical peers. Empirically we compare jazz musicians, classical musicians, and nonmusicians in a battery of psychophysical and EEG tasks. The psychophysical task (modified from Navarro Cebrian and Janata (2010)) entails perception and imagery of different musical scales, where participants' task is to judge whether the final pitch is too high, too low, or in tune. Jazz musicians show higher accuracy and a steeper psychometric function, suggesting heightened sensitivity to mistuned pitches given a tonal context. The EEG task (modified from Koelsch, Gunter, Friederici, and Schroger (2000)) compares expected, slightly unexpected, and highly unexpected chord progressions while participants rate the pleasantness of each chord progression. Given this explicit judgment task we see that the P300, an ERP component known to reflect explicit awareness and target processing, is enlarged during unexpected tonal harmonies for jazz musicians, and furthermore its amplitude is positively correlated with the length of musical training. Taken together, our central theme is that the process of improvisation requires heightened awareness of, and sensitivity to, tonal possibilities within a musical context, which allow the individual to generate novel sequences that are acceptable but original within the domain of jazz music.

I. BACKGROUND

A. Introduction

What characterizes the minds of exceptionally creative people, and how can we learn from them? Creativity is the ability to produce work that is novel (original, unexpected), high in quality, and appropriate (Sternberg, Lubart, Kaufman, & Pretz, 2005). To be deemed creative, a piece of work is defined relative to the field in which it lives, and thus must demonstrate some domain-specific knowledge on the part of its creator (Csikszentmihalyi, 1997). In the domain of music, improvisation is a form of spontaneous creative behavior that requires "novel combinations of ordinary mental processes" (Limb & Braun, 2008). Jazz musicians have been examined as a model of creativity due to the emphasis of improvisation in jazz musical performances (Limb & Braun, 2008; Pinho, de Manzano, Fransson, Eriksson, & Ullén, 2014). Functional MRI studies of jazz improvisation and other forms of spontaneous musical creativity generally show results in the frontal lobe, described in some reports as changes in functional connectivity and/or a tradeoff in activity between medial and dorsolateral prefrontal cortices (Berkowitz & Ansari, 2008; Donnay, Rankin, Lopez-Gonzalez, Jiradejvong, & Limb, 2014; Limb & Braun, 2008; Liu et al., 2012; Pinho et al., 2014). A thorough review of these and other neuroimaging results, however, suggested that these data patterns are somewhat inconclusive and sometimes in conflict between different studies, with evidence for creativity being supported by both activation and deactivation of the frontal lobe (Dietrich & Kanso, 2010).

B. The process of creativity

The complexity of fMRI results on creativity to date may arise from the diverse strategies that participants bring to bear when generating their creative output. The process of creativity by its very nature entails divergent thinking, which is commonly tested using divergent thinking tests (Runco, 1991), in which participants are given open-ended questions and tasked with generating as many responses as possible (Torrance, 1968). In contrast to most cognitive (convergent thinking) tests, divergent thinking tests yield no single correct answer. This poses a difficulty for the neuroscience of creativity, as it could be elusive to track down a single mental process of novel idea generation. One view of how novel ideas are generated comes from the theory of Blind Variation and Selective Retention (Campbell, 1960), in which organisms explore multiple candidates of possible ideas before selecting and implementing the most appropriate options. Time-sensitive measures of brain activity, when coupled with precise measures of each participant's given problem space and their resultant creative outputs, may test the hypothesis of exploration followed by selection in the creative process.

C. Expectation and sensitivity as domain-specific knowledge

While the BVSR theory provides a domain-general account for the cognitive processes necessary for divergent thinking, expertise and domain-specific experience may cut down the process of blind variation. A seasoned creator, such as a well-trained jazz musician, may shortcut the variation process by efficient use of domain-specific tools such as perceptual imagery and musical expectation, which are informed by long-term knowledge and sensitivity to statistically frequent and probable events in their environment (Huron, 2006).

Additional support for expectation and imagery as domain-specific knowledge comes from jazz pedagogy, in which the cognitive components that comprise teaching improvisation are viewed not as unitary, but as involving anticipation, use of learned repertoire, emotive communication, feedback, and flow (Biasutti, 2015). In particular, the state of anticipation involves the interface between expectation and perceptual imagery, both of which are widely studied with well-established paradigms in music perception and cognition in behavioral (psychophysical) tests, and in time-sensitive measures of electrical brain activity (Janata & Paroo, 2006; Koelsch et al., 2000).

D. The Present Research

Here we apply psychophysical, electrophysiological, and psychometric tools from music perception and cognition research to clarify our understanding of creativity. Specifically, we examine the roles of divergent thinking, expectation, and perceptual imagery in jazz musicians as a model of creativity, compared with non-improvising musicians and nonmusician control groups. A major advantage of the following tests is that they offer specific, controlled stimuli to couple with neural measures, thus cutting down the problem space for a more rigorous understanding of jazz improvisation as a domain of creativity.

II. AIMS

A. Overall hypothesis

Here we combine psychophysical measures of auditory imagination and perception (Janata & Paroo, 2006), behavioral and electrophysiological measures of musical expectation (Koelsch et al., 2000), and domain-general measures of divergent thinking (Torrance, 1968), to test the hypothesis that spontaneous musical creativity depends on 1) heightened perceptual awareness and more accurate mental imagery, 2) increased sensitivity to and awareness of unexpected events, and 3) heightened domain-general divergent thinking abilities. We test this hypothesis using jazz musicians as a model of spontaneous musical creativity, compared with non-improvising musicians and non-musician controls.

III. METHOD

A. Subjects

Subjects were recruited from Wesleyan University or the Hartt School of Music in exchange for monetary compensation or partial course credit. Subjects gave informed consent as approved by the Institutional Review Boards of Wesleyan University and Hartford Hospital.

Table 1. Subject characteristics in Jazz musician, Control (non-jazz) musician, and Non-musician groups

	N	% Female	Age (years) M(SD)	Pitch Discrim (Hz) M(SD)	Raw IQ (Shipley, 1940) M(SD)	Training Onset (years) M(SD)	Training Duration (years) M(SD)
Jazz Musicians	17	11.8	20.1 (1.5)	4.1 (3.2)	17.5 (1.8)	7.9 (2.8)	8.7 (3.4)
Control Musicians	16	50.0	22.4 (5.9)	5.0 (2.2)	16.5 (1.6)	8.8 (3.4)	10.0 (4.6)
Non-musicians	24	62.5	19.0 (1.1)	14.4 (13.6)	16.6 (2.3)	9.3 (2.8)	2.1 (2.1)

B. Scale Imagery Task

Participants listened to scales (either major, harmonic minor, or blues) and judged whether the last note was modified in pitch. The scales were played using Max/MSP and each note lasted for 250 ms with an inter-onset interval of 600 ms. All 12 keys were used randomly throughout the experiment with starting notes of F#3 (184.997 Hz) through F4 (349.228 Hz). There were 108 trials total with a block of 36 trials for each scale block, for which the order was rotated for each participant. The last note alterations were ±0, 25, 50, 75 and 100 cents and these alterations were randomized within each scale block.

There were two conditions: perception and imagery. For the perception condition participants were asked to judge whether the last note was higher, lower, or the same as the expected pitch. In the imagery condition the two second-to-last notes of the scale were silent, and participants were asked to imagine these two notes in the silent gap and still judge the last note. For each condition there was a practice round of 10 trials during which participants received feedback on the screen and the experimenter monitored their accuracy to make sure they understood the task.

Linear psychometric functions were fitted to yield the slope for imagery and perception conditions for each individual. General accuracy was compared in addition to slopes of psychometric functions between groups.

C. EEG Harmonic Expectation Task

Stimuli consisted of chord progressions that were either expected, slightly unexpected, or highly unexpected (figure 1). The participants were instructed to listen to each chord progression and rate their preference for it on a scale from 1-4, with the 1 being dislike and 4 being like. The trials were arranged in blocks of 60, and each participant completed at least 3 blocks (maximum 6 blocks). EEG was recorded using PyCorder software from a 64-channel BrainVision actiCHamp setup with electrodes corresponding to the international 10-20 EEG system. The recording was continuous with a raw sampling rate of 1000 Hz. EEG recording took place in a sound attenuated, electrically shielded booth.

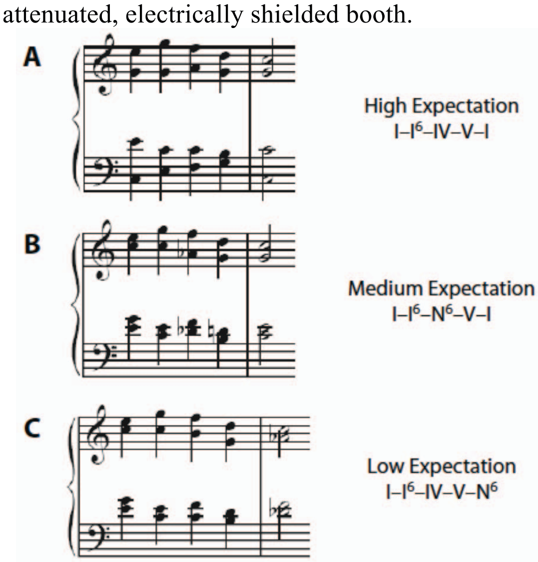


Figure 1. Example high, medium, and low expectation chord progressions.

Raw EEG data were imported to BrainVision Analyzer for analysis. Preprocessing included applying infinite impulse response filters with a low-pass cutoff of 30 Hz and a high-pass cutoff of .5 Hz. Raw data inspection was used to exclude data points with a higher gradient (>50uV/msec), high mins and max (>200uV), and extreme amplitudes (-200 to 200 uV). Ocular correction ICA was also done for each participant. The data were then segmented into chords and the trials were averaged and baseline corrected. We compared ERP traces for high, medium, and low expectation chords among the groups. We also plotted difference waves for medium minus high expectation and for low minus high expectation. Peaks for each subject were then exported from BrainVision Analyzer and analysed separately in SPSS.

D. Divergent Thinking Task

Participants responded to 6 open ended prompts for three minutes (Torrance, 1968). Participants were told that the task was a measure of general creativity and that they should try to give as many answers as they could. Participants' responses were coded for fluency and originality. Fluency was calculated as the number of unique responses. Responses from 16 control participants (nonmusicians) were used to create a baseline for originality. The participants were then scored for originality with unique responses receiving 3 points, responses that occurred once in the baseline receiving 2 points, and responses that occurred twice in the baseline receiving 1 point.

IV. RESULTS

A. Scale Imagery Task

A mixed factor ANOVA on the dependent variable of accuracy with the between-subjects factor of group (Jazz musicians, Non-jazz musicians, Non-musicians) and the within-subjects factor of task (perception vs. imagery) showed significant main effects of group ($F(2,33) = 22.8$, $p < .001$) and task ($F(1,33) = 21.0$, $p < .001$) but no task-by-group interaction ($F(2,33) = .65$, n.s.). A mixed-factor ANOVA on the dependent variable of slope, with the between-subjects factor of group and the within-subjects factor of task, showed a main effect of group ($F(2,33) = 11.2$, $p < .001$) and a main effect of task ($F(1,33) = 23.3$, $p < .001$) but no significant interaction between group and task ($F(2,33) = .027$, n.s.). These results, also shown in Figure 2, confirm that jazz and non-jazz musicians are more accurate at detecting mistuned scales in both perception and imagery.

B. EEG Harmonic Expectation Task

1) Behavioral Data. A mixed factor ANOVA with the within-subjects factor of expectation (high, medium, low) and the between-subjects factor of group (Jazz, Non-jazz including musicians and non-musicians) showed a main effect of expectation ($F(2,21) = 13.6$, $p < .001$) on preference ratings, as well as an interaction between expectation and group ($F(2,21) = 5.3$, $p = .014$). Preference ratings showed that jazz musicians prefer the medium expectation condition ($t(10) = 3.5$, $p = .005$) as compared to the non-jazz subjects (including non-jazz musicians and non-musicians) who prefer the high expectancy chords. While all groups showed lowest preference ratings for the low expectation condition, ratings for the low expectation condition was higher for the jazz musicians ($t(22) = 2.2$, $p = .03$), suggesting higher tolerance for unexpected events among the jazz group. This provides support for the notion that affect is aroused in music by slight violations of expectations (Meyer, 1956).

2) ERP Data. A mixed factor ANOVA on the dependent variable of ERP amplitude during the last chord, with the between-subjects factor of group and the within-subjects factor of expectancy (low vs. high) showed a significant main effect of expectancy and a significant interaction between expectancy and group for electrodes P2, P4, and PO4 between 410-480 ms and F8 and FT8 between 220 and 260 ms (See Table 2 for and F and p values).

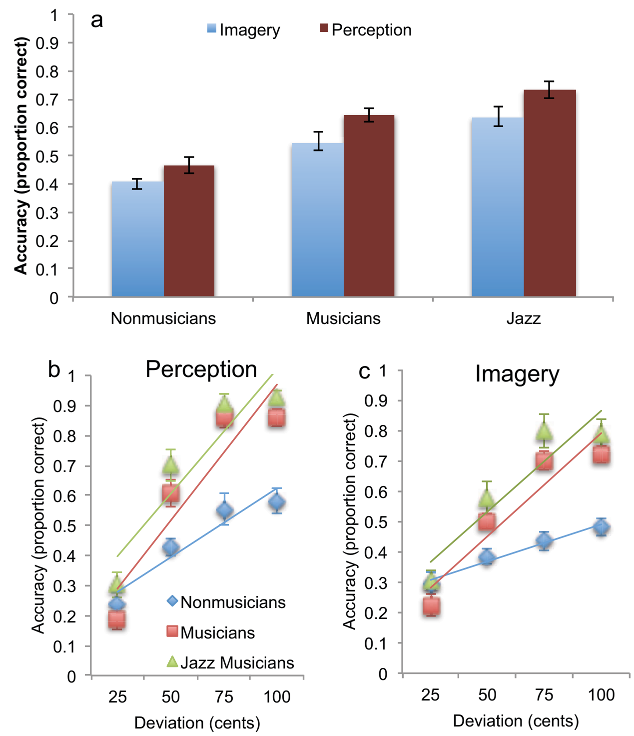


Figure 2. a. Accuracy on scale perception and imagery. b. Psychometric functions for scale perception task. c. Psychometric functions for scale imagery task.

Table 2. Analysis of the ERP data via mixed factor ANOVA

Electrode	Time (ms)	Main Effect of Expectancy	Expectancy x Group Interaction
P2	410-480	F(1,33) = 29.3, p<.001	F(1,33)=4.1, p<.05
P4	410-480	F(1,33)= 28.1, p<.001	F(1,33)=4.3, p<.05)
PO4	410-480	F(1,33)=22.4, p<.001	F(1,33)=4.4, p<.05
F8	220-260	F(1,33)=8.1, p<.01	F(1,33)=3.7, p<.05
FT8	220-260	F(1,3)=6.4 p<.05	F(1,33)=4.6, p<.05

C. Divergent Thinking Task

One-way ANOVAs on the dependent variable of fluency, with the factor of group (Jazz musicians, Non-jazz musicians, Non-musicians) showed significant main effects for questions 3 ($F(2,31) = 8.7$, $p < .01$), 4 ($F(2,31) = 7.3$, $p < .01$), 5 ($F(2,31) = 6.6$, $p < .01$), and 6 ($F(2,31) = 4.4$, $p < .05$). One-way ANOVAs on the dependent variable of originality showed significant main effects for questions 2 ($F(2,31) = 3.6$, p

= .04), 3 (F(2,31) = 13.1, p <.001), 4 (F(2,31) = 12.0, p <.001), 5 (F(2,31) = 8.7, p =.001), and 6 (F(2,31) = 5.5, p =.01). The jazz musicians scored the highest out of the three groups on questions 4, 5, and 6 for fluency and questions 3, 4, 5, and 6 for originality followed by musicians and then non-musicians.

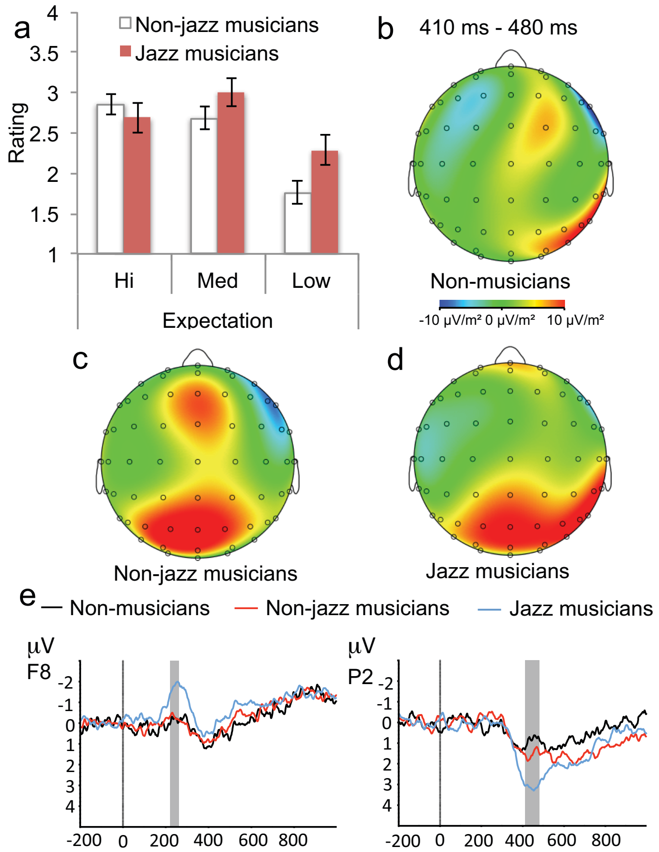


Figure 3. a. Behavioral ratings for high, medium, and low-expectation chord progressions. b-d. Topos plots of the 410-480 ms time window after the onset of the last chord in the low-expectation condition in non-musicians (b), non-jazz musicians (c), and jazz musicians (d). All color scales range from -10 to +10 µV. e. Difference waves (low minus high expectation) for last chords for right frontal (F8) and right parietal (P2) electrodes, showing enhanced early negativity and late positivity in jazz musicians.

V. DISCUSSION

Results from psychophysical, electrophysiological, and psychometric tasks converge to show superior auditory imagery and scale perception, heightened sensitivity to expectation, and higher domain-general creativity in Jazz musicians. These results provide support for the use of Jazz musicians as a model for creativity.

Psychometric functions show steeper slopes for both groups of musicians compared to non-musicians, suggesting that musical training in general enhances perceptual and imagery sensitivity. Auditory imagery is an important skill for musical performers of all genres, as musicians often have to be able to imagine an upcoming note or chord before it happens in order to craft their performance accordingly. Notably, these effects are observed despite similar baseline levels of performance on a pure tone pitch discrimination task (Table 1). By providing a musical context, the scale perception and imagery tasks assess a more central, memory-dependent strategy, and are thus more dependent on training compared to the lower-level pitch discrimination task.

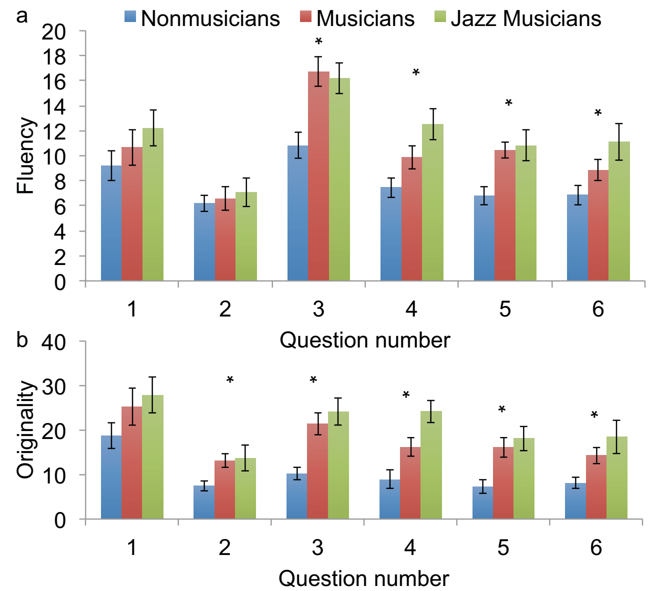


Figure 4. Results from divergent thinking task scored for (a) fluency and (b) originality. * = p < .05 (one-way ANOVAs).

Behavioral results for the EEG chord progression ratings task differ from previous studies (Loui & Wessel, 2007), which showed that musicians and non-musicians have similar preferences. Here, Jazz musicians' preference for the medium expectation chords as opposed to the high expectations chords may be due to the nature of Jazz where the rules are more free and meant to be broken in some instances of improvisation. The significantly higher rating of the low expectancy chords also suggests that Jazz musicians are more tolerant to chords that sound out of place. This may be explained by the experimental nature of jazz improvisation, where it is customary to embellish performances by violating expectations. In contrast to contemporary classical musical training, jazz improvisers are encouraged to play notes and chords that seem out of place, as many Jazz musicians use chords that seem out of place as a transition to a new tonal landscape or musical idea.

ERP results show interactions between group and expectation in right-hemisphere electrodes (table 2). This suggests that the different groups respond to the unexpected chords differently. Difference waves clearly show that Jazz musicians have larger Early Right Anterior Negativity (ERAN) and P3 (figure 3e). Interestingly, both the early negativity and the P3 components are right-lateralized in Jazz musicians, with the P3 especially more in the right posterior electrodes for the Jazz musicians and the left posterior and frontocentral electrodes for the Non-jazz musicians (figure 3b-d). This could be supported by neuropsychological findings on hemispheric asymmetry (Ivry & Robertson, 1997) as well as fMRI work on creative musical sequence generation (Villarreal et al., 2013), where the right hemisphere is shown to subserve holistic perception and creative thought. Together these results indicate that Jazz musicians have an enhanced sensitivity to harmonic expectation compared to the Non-jazz musician group and the Non-musician group. These findings are in contrast to no differences in low-level perceptual abilities and age of onset or number of years of musical training between non-jazz musicians and jazz musicians (as shown in the control pitch discrimination tasks), and no differences in IQ among the three groups.

The Jazz musicians' high performance on the DTT on 4/6 questions compared to Non-jazz musicians and Non-musicians indicates that Jazz musicians have a general advantage in creativity that transcends the domain of music. We believe that questions 1 and 2 may not have captured significant differences among the three groups because 1) the questions were especially ambiguous and resulted in extremely divergent answers, and 2) possible order effects as participants might have needed time to engage themselves fully in the task of divergent thinking. Nevertheless, the differences between the Jazz musicians and other two groups for the rest of the questions is striking given that there are no differences in IQ among the three groups.

VI. CONCLUSION

Jazz musicians in our present sample scored higher on domain-general creativity tasks. Psychophysical and electrophysiological measures suggest that they also possess heightened perceptual awareness of and sensitivity to unexpected events within a musical context. Taken together, results from domain-specific as well as domain-general tasks suggest that creativity entails being open to unexpected events within one's domain, as well as being more fluent and original in idea generation. The present results validate the use of jazz improvisation as a model system for understanding creativity, and further suggest that systematic violations of domain-specific expectations may provide a time-sensitive measure of the rapid and flexible real-time creative process.

ACKNOWLEDGMENT

Supported by the Imagination Institute. We thank all our participants.

REFERENCES

Berkowitz, A. L., & Ansari, D. (2008). Generation of novel motor sequences: the neural correlates of musical improvisation. *Neuroimage, 41*(2), 535-543.

Biasutti, M. (2015). Pedagogical applications of the cognitive research on music improvisation. *Frontiers in Psychology, 6*.

Campbell, D. T. (1960). Blind variation and selective retentions in creative thought as in other knowledge processes. *Psychological Review, 67*(6), 380-400.

Csikszentmihalyi, M. (1996). *Creativity: flow and the psychology of discovery and invention*. New York: HarperCollinsPublishers.

Dietrich, A., & Kanso, R. (2010). A review of EEG, ERP, and neuroimaging studies of creativity and insight. *Psychol Bull, 136*(5), 822-848.

Donnay, G. F., Rankin, S. K., Lopez-Gonzalez, M., Jiradejvong, P., & Limb, C. J. (2014). Neural Substrates of Interactive Musical Improvisation: An fMRI Study of 'Trading Fours' in Jazz. *PLoS ONE, 9*(2), e88665.

Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation* (1 ed. Vol. 1). Cambridge, MA: MIT Press.

Ivry, R. B., & Robertson, L. (1997). *The Two Sides of Perception*: MIT Press.

Janata, P., & Paroo, K. (2006). Acuity of auditory images in pitch and time. *Percept Psychophys, 68*(5), 829-844.

Koelsch, S., Gunter, T., Friederici, A. D., & Schroger, E. (2000). Brain indices of music processing: "nonmusicians" are musical. *J Cogn Neurosci, 12*(3), 520-541.

Limb, C. J., & Braun, A. R. (2008). Neural substrates of spontaneous musical performance: an FMRI study of jazz improvisation. *PLoS One, 3*(2), e1679.

Liu, S., Chow, H. M., Xu, Y., Erkkinen, M. G., Swett, K. E., Eagle, M. W., . . . Braun, A. R. (2012). Neural correlates of lyrical improvisation: an FMRI study of freestyle rap. *Sci Rep, 2*, 834.

Loui, P., & Wessel, D. (2007). Harmonic expectation and affect in Western music: Effects of attention and training. *Perception & Psychophysics, 69*(7), 1084-1092.

Meyer, L. (1956). *Emotion and Meaning in Music*: U of Chicago Press.

Navarro Cebrian, A., & Janata, P. (2010). Electrophysiological correlates of accurate mental image formation in auditory perception and imagery tasks. *Brain Res, 1342*, 39-54.

Pinho, A. L., de Manzano, Ö., Fransson, P., Eriksson, H., & Ullén, F. (2014). Connecting to create: Expertise in musical improvisation is associated with increased functional connectivity between premotor and prefrontal areas. *The Journal of Neuroscience, 34*(18), 6156-6163.

Runco, M. A. (1991). *Divergent thinking*. Westport, CT US: Ablex Publishing.

Shipley, W. C. (1940). A self-administering scale for measuring intellectual impairment and deterioration. *Journal of Psychology, 9*, 371-377.

Sternberg, R. J., Lubart, T. I., Kaufman, J. C., & Pretz, J. E. (2005). Creativity. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (Vol. 137, pp. 351-369): Cambridge University Press Cambridge.

Torrance, E. P. (1968). Examples and rationales of test tasks for assessing creative abilities *Journal of Creative Behavior, 2*(3), 165-178.

Villarreal, M. F., Cerquetti, D., Caruso, S., Schwarcz Lopez Aranguren, V., Gerschcovich, E. R., Frega, A. L., & Leiguarda, R. C. (2013). Neural Correlates of Musical Creativity: Differences between High and Low Creative Subjects. *PLoS One, 8*(9), e75427.

All You Need to Do is Ask? The Exhortation to Be Creative Improves Creative Performance More for Non-Expert than Expert Jazz Musicians

David S. Rosen[1], Youngmoo E. Kim[2], Daniel Mirman[1] and John Kounios[1]
[1]Psychology, Drexel University, Philadelphia, PA, USA
[2]Electrical and Computer Engineering, Drexel University, Philadelphia, PA, USA

Current creativity research continues to disagree about the nature of creative thought, specifically, whether it is primarily based on Type-1 (automatic, associative) or Type-2 (executive, controlled) cognitive processes. We proposed that Type-1 and Type-2 processes make differential contributions to creative production depending on domain expertise and situational factors such as task instructions. We tested this hypothesis on jazz pianists who were instructed to improvise to a novel chord sequence and rhythm accompaniment. Using a staggered-baseline design, participants were introduced to explicit creativity instructions, which, via goal activation, is thought to prompt the musicians to engage Type-2 processes. Jazz experts rated all performances with high degrees of inter-rater reliability. Multilevel regression models revealed a significant interaction between Expertise and Instructions (standard / 'be creative') when predicting musicians' improvisation ratings. Overall, performances by more experienced pianists were rated as superior and creativity instructions resulted in higher ratings. However, creativity instructions benefited less-experienced pianists more than it did for more-experienced pianists, suggesting differential contributions by Type-1 and Type-2 processes depending on situational factors and individual differences. These findings indicate that training and experience afford jazz pianists the ability to develop efficient creative processes, relying more on implicit, unconscious cognitive systems than novices. Since explicit "be creative" instructions are a challenging goal that occupies the conscious mind, they interfere with the optimal creative processes of expert jazz musicians; however, for less experienced musicians, consciously attending to a creative goal can help them avoid cognitive fixation, assisting musicians to alter their performances in ways that facilitate more creativity.

Negentropy: A compositional and perceptual study using variable form

Jason C. Rosenberg[1]
[1]Center for New Music, San Francisco, CA, USA

Variable form compositions permit multiple orderings of its sections and can thus be used to study preferences in musical form. This study uses a contemporary musical composition with variable form (created by the author, a composer-researcher) to investigate 1) perceived similarity of musical material and 2) preferences when creating large-scale forms. The composition, titled *Negentropy*, is a stand-alone concert piece for solo flute, intended to double as a music perception task. It is divided into 9 sections (between 25 and 55 seconds each) that can be variably sequenced, resulting in just under a million possible unique sequences. Participants (7 professional composers and 7 non-musicians) characterized each of the 9 sections, then sorted the sections into groups based on musical similarity, and then ordered the sections (omitting up to 3) into their own "compositions".

Preliminary results reveal that in the sorting task divergent sections were reliably placed into the same groups, suggesting that listeners prioritized surface similarities in their classifications rather than form-bearing dimensions, such as structural organization. Data from the sequencing task reveal an unexpected agreement between composers' and non-musicians' sequences. 71% of participants chose a sequence with a symmetrical compositional structure consisting of tender, contemplative outer sections and energetic, rhythmic inner sections. That is, composers and non-musicians alike preferred an arch form ordering. It was anticipated that composers would choose more unconventional sequences. In contrast, the two flautists that have performed the piece thus far chose non-traditional forms, thereby implying that increased familiarity with the musical material, rather than musical experience, may result in more daring structural decisions. In sum, the *Negentropy* project contributes to our understanding of preference in large-scale musical form, and presents an integration of scientific inquiry with contemporary musical composition.

A method for identifying motor pattern boundaries in jazz piano improvisations

Martin Norgaard[1]
[1]School of Music, Georgia State University, Atlanta, Georgia, USA

According to Pressing's (1988) prominent theoretical framework, musical improvisations consist of concatenated motor patterns. Accordingly, boundaries between these patterns should exist in extant improvisations. Schmidt (1975) defined Generalized Motor Patterns (GMPs) as an abstract movement structure in which the proportions of relative timing and relative force are constant. He later identified the boundary between two GMPs using correlations among times of kinematic landmarks within a longer lever movement. Here I adapted this idea to musical improvisation. I analyzed a corpus of 1000 choruses of jazz blues improvisations played by 25 advanced jazz pianists on a weighted midi keyboard. Thirty-six different five-note patterns were identified with identical rhythm and pitch content that occurred 10 times or more in the corpus. I then ran correlations of relative force and relative timing between the first note and each of the following six notes to determine if subsequent notes were part of the same motor pattern. Six of the 36 patterns showed strong correlations in both force and timing indicating that most of the 25 participants used stored motor patterns to execute these patterns. Interestingly, all of these patterns were either ascending or descending eighth note chromatic lines. Future research could use this method to analyze output from one participant, which could identify the player's idiosyncratic patterns. These results lend support to Pressing's (1988) model in which players insert stored motor patterns during musical improvisation.

The meaning of making – Mapping strategies in music composition

Peter Falthin

Abstract—One way to think of creative processes is as recontextualizations of perceptions and conceptions of reality. Impressions and ideas are seen from new perspectives and connected in new ways before entered into a new context in a different form, which may or may not include shifts in modality or form of representation. This study is about how composition students give musical expression to extra-musical phenomena and how they relate their musical thinking to other forms of representation. It involves studying what mapping strategies the student composers develop in order to establish relationships between different forms of representation, but also to study the meaning making processes in both the analysis and synthesis phase of the restructuring of concepts.

The how-questions imply a qualitative approach and method. Data comprise a wide variety of sketch material, as well as scores, performances and recordings of the finalized compositions, and in-depth interviews with the student composers in relation to these materials. In all the studied cases, composition process began with extramusical considerations in the form of narratives, imagery or some kind of physical phenomena (e.g. geometrical concepts, acoustical phenomena and tactile qualities). Typically there would appear several creative processes in different modalities converging into musical form along the composition process. Results suggest that these students intend their music to represent extramusical phenomena and concepts in as far as they take that as points of departure for developing compositional concepts, but also for shaping musical expression. To a varying degree, these extramusical considerations are meant to be conveyed in the music.

***Keywords*—Composition, Mapping strategies, Musical meaning making, Symbolic representation**

I. Introduction

Music is one of the many ways humans try to make sense of the world; to comprehend, form, and express aspects of meaning. Musical meaning making can be seen in a multitude of ways. In focus for this study is representational and symbolic meaning making in the course of the composition process and as part of musical expression.

The purpose of this study is to understand the complexity of the esthetic aims and developments of student composers at university college, particularly in terms of symbolic representation, how and to what ends it is applied, and the nature and role of semiotic process in musical meaning making. These meaning making processes emerge in the course of composition and operate on several levels. In the projects in this study, these meaning making processes typically serve to connect aspects of thinking about worldly or extramusical matters to musical structure, for inspiration to render seed material, for structuring musical form and for expressive purposes.

The research question is: How do composers use mapping strategies to integrate source material intor their compositions, and to convey musical meaning to musicians and listeners?

P. Falthin is with the Royal College of Music in Stockholm, Box 27711, 11591 Stockholm, Sweden (phone: +46-704715220 e-mail: peter.falthin@kmh.se).

II. Approach and Methods

The study was inspired by ethnological methods in the sense that it was longitudinal and strived towards an ecological understanding of the objects, and functions of the meaning making processes[1]. Three composers took part in the study. Composition processes often stretch out in time and pass through different stages and consequently learning and development in composition are long term endeavors, which is a reason this study has stretched over seven years, from February 2010 to May 2016. During that period the participants have developed and refined the techniques and methods for composing, and esthetic principles displayed in this study as well as their musical knowledge in a broader sense. At the beginning the participants were undergraduate students in music composition at a university college, but since then they have all finished their composition studies and are now professional composers. One of them has also taken a masters degree in music composition and another is about to finish a master's degree in mathematics.

Methods for data collection comprised interviews, analyses of scores, sketch materials, performances and recordings. The results presented in this paper constitute but a fraction of the total data material, and were selected to represent the particular foci of this paper – mapping strategies and how meaning emerges from activities of making.

The musical analyses, including analysis of the compositional techniques used, were to a large degree done in dialogue between composer and researcher in the course of the interviews. The outcome of these analyses were subjected to semiotic analysis and structured along three different strands: 1. Aspects of extra musical meanings used for inspirational purpose or to generate seed material for composition. 2. Sketch materials or models to structure musical form or regulate construction principles. 3. Symbolic representations to make extramusical references as an integral part of the musical expression.

III. Musical topology

Relations berween spatial notions and music can be approached from different angles. That physical gesture is inherent in musical gesture was a point of departure for Smalley when developing spectromorphology, as a technique for notating gestural qualities of sonic events[2]. Smalley mentions four orders of surrogacy that can be engaged whenever a physical gesture is propagated in sound, going from concrete to abstract relations, and discusses the problem of source-bonding as limiting to extrinsic meaning making in music.

Tim has developed techniques for graphical sketching of musical ideas, based on an experienced strong connection between spatiality and music. The sketches employ symbolic representation with a certain degree of conceptual flexibility. The symbols have a kind of nodal structure that is context dependent. According to Tim, items in the drawings represent topoi in the composer's mind that correspond to certain musical expressions. Sometimes the sketches imply a time axis, other times their function is more geographical, like mind maps over musical ideas and their contexts. In order to elaborate the technique and esthetic principles, Tim practiced this kind of sketching separate from compositional work, as a discipline in itself (fig. 1) "to develop a language of musical drawing" as he put it. Sketching in this way offers a kind of abstraction; a meta-level for organizing musical thinking with an option to infer quite specific musical ideas prior to designing the musical activities to shape the musical expression.

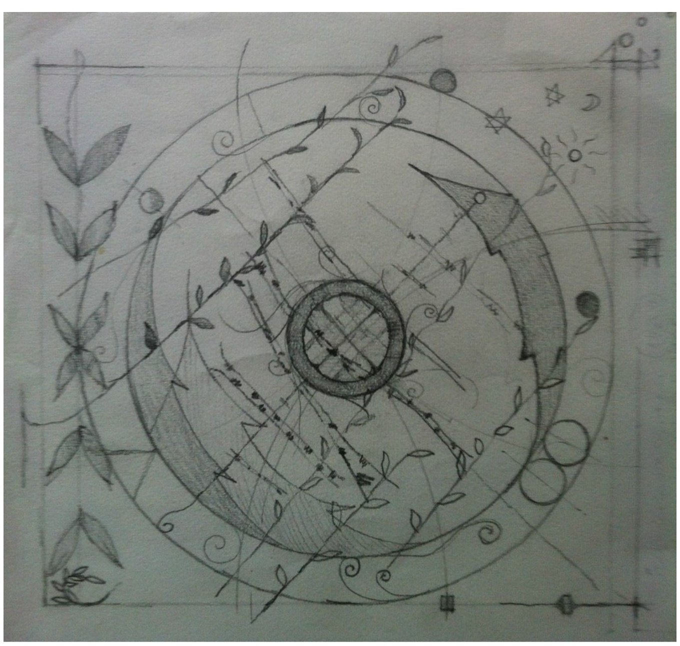


Fig. 1 A study to develop graphical sketching of music.

Tim has developed the sketching into notation technique, often used in combination with elements of traditional notation. An important feature in the early phases of this was a grid to guide the musician in the awareness of different degrees of presence in the music by means of a simple spatial model. The degree of presence in the music could concern such different aspects as the freedom of improvisation in relation to the score, or a sense of abstraction of being in the world.

IV. Generative structure

Generative structures are commonplace in music composition and typically engage computerized algorithms that mimic some aspect(s) of creation in nature, often inspired by evolutionary theories[3][4]. Ultimately it addresses the question 'what is life?' and problems of the existence of a free will. In a series of composition Vincent has dealth with a different approach to generative structures for in music composition, wherein a set of instructions enrols the musicians as generators of musical events and gestures. The overarching structure of the instructions together with the actions of the musicians provide the local and global context; the form of the piece.

Vincent has provided very few sketches within the frame of the study. Instead his mapping strategies go right into the scores that typically focus musical activity rather than convey an image of a musical idea. Some of his pieces have a game-like structure where the actions of the musician have a major influence on the musical structure. This is true in a piece for any instrument, which is setup like a maze where the musician is faced with a series of choices. The choices made, lead to a path that brings certain obligations and then to new forks in the path.

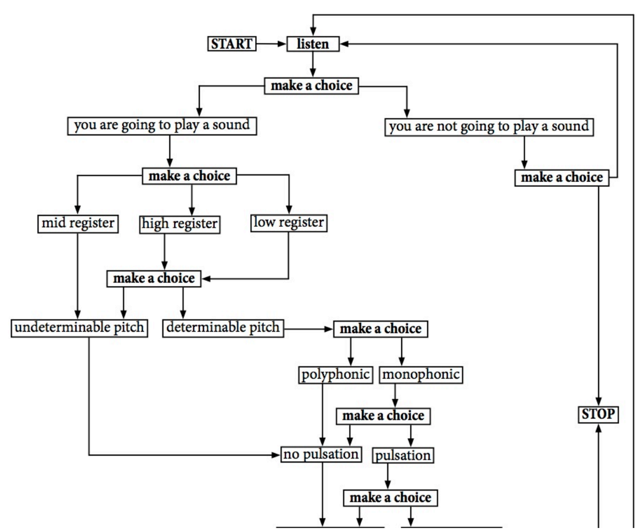


Fig. 2 Excerpt from the piece Choices and Obligations.

The design resembles that of geographical maps, but in addition to displaying possible pathways it provides a framework for musical improvisation.

V. Synthesis and Communication

Liv's compositional process sparks from finding commonalities or correspondence between seemingly disparate phenomena, sometimes of musical origin and sometimes from extra-musical sources. The different source materials are processed through each other to render materials for the composition and the compositional process, often in several different dimensions, e.g. time proportions, harmonic structure, instrumentation (sometimes even redesigning of instruments) and design of the composition process. A literary dimension is integral to Liv's approach to musical meaning making and compositional work. Mapping strategies in Liv's work always include transformation; the different materials actually change and new meanings emerge.

In a series of works connected through internal heritage, where the resynthesis of source materials rendered stuff to become a seed for the next piece and so on, there was a piece for violin and electro acoustics that employed complex sketching in multiple layers. One set of sketches concerning rhythmical proportions and gesture, involved calculations for

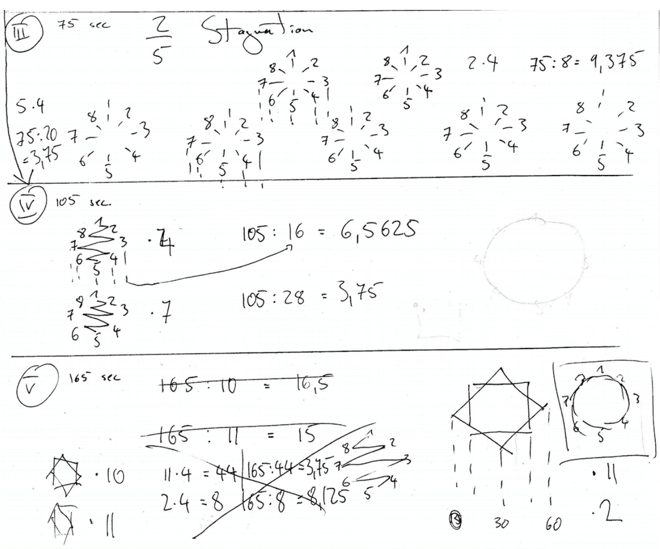


Fig. 3 Sketch guiding spatialisation and composition process.

phrase length and metric development. Another set of sketches regulated the spatializations of the electro acoustic part. There were sketches concerning development and application of a folk tune used as source material and sketches to deal with the distribution of playing techniques and a drawing of the disposition and proportions of the parts in the large scale form. But lastly there was a drawing to steer to what extent some of the other drawings and sketches should be followed in the compositions process. This way, there was a conflict introduced between the strictly regulated parametrical schemata and randomly assigned instances of free will.

The mapping strategies here involves both symbolic representation, mapping aspects of extrinsic meaning to musical entities, and an intricate web of mapping different aspects of the composition process together.

References

[1] J. Blacking, *How musical is man?* Seattle ; London: University of Washington Press, 1973.

[2] D. Smalley, "Spectromorphology: explaining sound-shapes," *Organised Sound*, vol. 2, no. 2, pp. 107–126, Aug. 1997.

[3] P. Dahlstedt, "Between Materials and Ideas: A Process-Based Spatial Model of Artistic Creativity," in *Computers and Creativity*, vol. 2012, McCormack, D'Inverno, 2012.

[4] M. A. Boden, *The creative mind: myths and mechanisms*, 2nd ed. London ; New York: Routledge, 2004.

Study of Instrument Combination Patterns in Orchestral Scores from 1701 to 2000

Song Hui Chon,*[1] Dana DeVlieger,*[2] David Huron*[3]

*School of Music, Ohio State University, U.S.A.
[1]chon.21@osu.edu, [2]devlieger.1@osu.edu, [3]huron.1@osu.edu

ABSTRACT

Orchestration is the art of combining instruments to realize a composer's sonic vision. Although some famous treatises exist, instrumentation patterns have rarely been studied, especially in an empirical and longitudinal way. We present an exploratory study of instrumentation patterns found in 180 orchestral scores composed between 1701 and 2000. Our results broadly agree with qualitative observations made by music scholars: the number of parts for orchestral music increased; more diverse instruments were employed, offering a greater range of pitch and timbre. A hierarchical cluster analysis of pairing patterns of 22 typical orchestral instruments found three clusters, suggesting three different compositional functions.

I. INTRODUCTION

There exist a number of classic orchestration treatises (e.g., Berlioz, 1855; Widor, 1904; Rimsky-Korsakov, 1912/1964; Koechlin, 1954-9; Piston, 1955; Adler, 2002). These books tend to offer a survey of acoustic properties of instruments, often according to the instrument family, such as pitch range and timbral characteristics in distinctive registers. They also provide recommended or disparaged instrument combinations with little systematic discussion regarding why certain examples work whereas others do not. There is surprisingly little scientific literature on the implicit principles that might guide composers' choice of instrument combinations. This lack of existing knowledge motivated us to survey the instrumentation patterns found in orchestral music. As a first step, we took an exploratory rather than hypothesis-driven approach.

In one of few empirical studies of orchestration, Johnson (2011) examined the instrument combination patterns found in the orchestral music of the Romantic period. He coded the instruments sounding at a randomly selected moment (*sonority* hereafter) from each of 50 orchestral scores and performed hierarchical clustering based on the combination patterns. The results revealed three clusters, which he dubbed *Standard*, *Power*, and *Color* (SPC) groups. The *Standard* cluster consisted of violin, viola, cello, contrabass, flute, oboe, clarinet, and bassoon. The *Power* cluster comprised piccolo, trumpet, trombone, French horn, tuba, and timpani. The *Color* cluster included English horn, cornet, harp, bass clarinet, and contrabassoon. He suggested that the three largest instrument clusters reflect different compositional goals: *Standard* instruments represent a sort of default instrumentation; *Power* instruments are used to convey energy, forcefulness, or potency; and *Color* instruments are typically deployed to produce novel, exotic, or idiosyncratic effects.

Inspired by Johnson's study, we explore the patterns of orchestration in a more longitudinal setting within a historical context, focusing on changes between 1701 and 2000. We examine the rise and fall of various orchestral instruments in terms of *instrumentation presence* (how often an instrument is included in an orchestral work) and *instrument usage* (how often an instrument is actually sounding in a piece).

II. METHOD

We first divided the 300 years into six 50-year epochs. Then in each epoch, 30 orchestral compositions were selected. In selecting the scores, we made use of the *Gramophone Classical Music Guide* (Jolly, 2010), *The Essential Canon of Orchestral Works* (Dubal, 2002), and the *list of composers* on Wikipedia. The composers were grouped into three tiers: Tier-1 for those listed in both the *Gramophone Guide* and the *Essential Canon*; Tier-2 for those listed in any one of the two books; Tier-3 for those listed only on the Wikipedia page. For each epoch, we started with a list of composers who might belong to the epoch, based on their birth and death years. We would sample one work per composer starting from Tier-1, based on the year of composition (or publication). If more works needed to be sampled, we would move to Tier-2 composers then to Tier-3. Sampling continued until 30 works were selected, or the list of composers was exhausted, in which case we would restart from Tier-1 composers. Sometimes a composer wrote works in more than one epoch. In such case, this composer's works would be sampled in both applicable epochs. In any case no more than three works by a single composer were sampled across all epochs. In terms of scores, we used those available at the OSU music library and the International Music Score Library Project (IMSLP).

From each score, 10 sonorities were randomly sampled and coded for instrumentation, pitch, dynamics, tempo, and year of composition. In terms of pitch, the concert pitch was coded. The dynamic level was identified according to the most immediately preceding dynamic marking, which is often an Italian abbreviation such as *mp* or *ff*. The markings of crescendo or decrescendo that modifies dynamics was also coded as "+" or "–". Hence, for example, "*p* followed by decrescendo" would be coded as "*p*–". Trills and ornaments were coded as the pitch of the main note. Articulation marks such as *staccato* or *sforzando* were ignored.

Tempo terms were coded as notated in the score by the composer. The Italian tempo terms (such as *Andante*) have become more or less a norm in Western classical music. However, many terms convolved style or genre terms with tempo—as in the term *gigue*. Although these terms might have implied specific tempos at the time they were used, from the modern vantage point the specific tempos implied are ambiguous. We also observed an increase of non-Italian tempo terms and/or tempo designation with metronome markings (such as ♩ = 72) through the years.

Over the 300 years some instruments went through stages of development. For the purposes of this study a number of instruments were deemed equivalent *a priori*. For example, the baroque transverse flute was deemed the same as the modern Boehme metal flute (hence both were coded as flute).

Additional questions of equivalence arise with regard to keyboard instruments. In some cases—such as with J.S. Bach's concertos—keyboard instrumentation may be ambiguous with a simple part name of *Clavier* (keyboard). In practice, this may have been harpsichord, fortepiano, or organ. Since our aim is to track changes over history, we resolved to interpret the keyboard instrument according to the practices at the time of composition. Thus, we elected to code works composed during or before 1780 as for harpsichord, and those after for piano, following the keyboard music culture around Mozart's time. Table 1 identifies instrumental equivalences for this study. The coding of continuo parts also raised special considerations. Unless the instrumentation was specified (such as bassoon, cello, and harpsichord parts designated as continuo as a group), the continuo part was encoded as simply continuo. No effort was made to realize figured bass notations. Instead, figured bass was simply coded as the bass pitch.

Table 1. List of historic instruments and their modern equivalences

Historic instruments	Modern(equivalent) instruments
Chalumeau, Basset Horn	Clarinet
Transverse flute, Flauto d'amore	Flute
Oboe d'amore, Oboe da silva, Oboe da caccia	Oboe
Hunting horn, Corno da caccia	French Horn
Baritone horn, Euphonium	Trombone
Viola da gamba	Cello
Violone	Contrabass
'Keyboard' or 'Clavier' before 1780	Harpsichord
'Keyboard' or 'Clavier' after 1780	Piano

III. RESULTS

The resulting sample set included a total of 165 parts employing 91 unique instruments. Many of these parts were of common orchestral instruments (such as violin, flute, and trumpet), whereas some parts disappeared in history (such as continuo). Some other instruments appeared more often with advancing years, including percussion and wind instruments often exhibiting extended pitch and timbre ranges (such as alto flute or contrabassoon). Analyses were carried out using instrumentation presence and instrument usage.

First, we might want to study the change in ensemble size over time. Due to the practice of sometimes having more than one instrument playing the same part, we do not have data for the total number of instruments required to perform a given piece of music. However, we can still examine the total number of parts notated in the score, which is likely to be proportional to the ensemble size. Figure 1 shows the total number of parts for the 180 orchestral scores in our sample. Each dot represents a score, with the corresponding composition (or publication) year on the horizontal axis and the total number of parts specified in the piece on the vertical axis. We can observe a constant growth of the number of parts until around the mid-20th century, after which time smaller-scale compositions started to reappear.

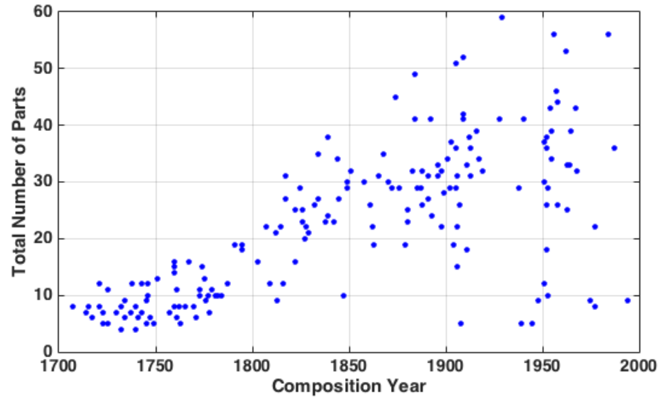


Figure 1. Scatterplot of the total number of parts specified in an orchestral score and its composition year

Figure 2 shows the instrumentation presence (top) and instrument usage (bottom) of string instruments. Even though cello and contrabass were not typical in early 18th century orchestras, they quickly became a core member of an orchestra in the 19th century and onward. The usage patterns of these instruments show a decreasing pattern through the years.

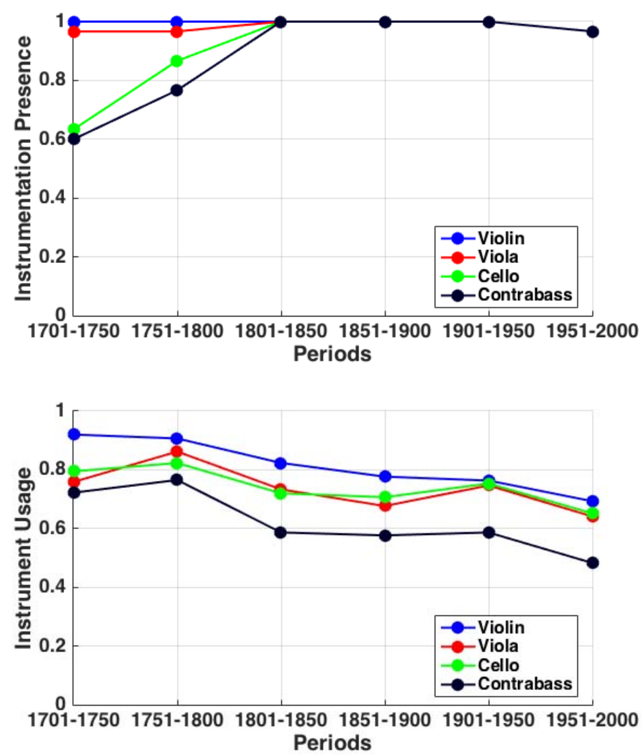


Figure 2. Instrumentation presence (top) and instrument usage (bottom) of string instruments

In Figure 3, we see that the presence of four instruments—oboe, flute, bassoon, and clarinet—is noticeably higher than others. Oboe, flute, bassoon, and clarinet were available earlier in history, which probably helped establish their presence in orchestras, whereas the other four instruments were introduced later presumably to provide a wider range of pitch and timbre. This separation of two groups can be still observed in the usage pattern, although not as clearly.

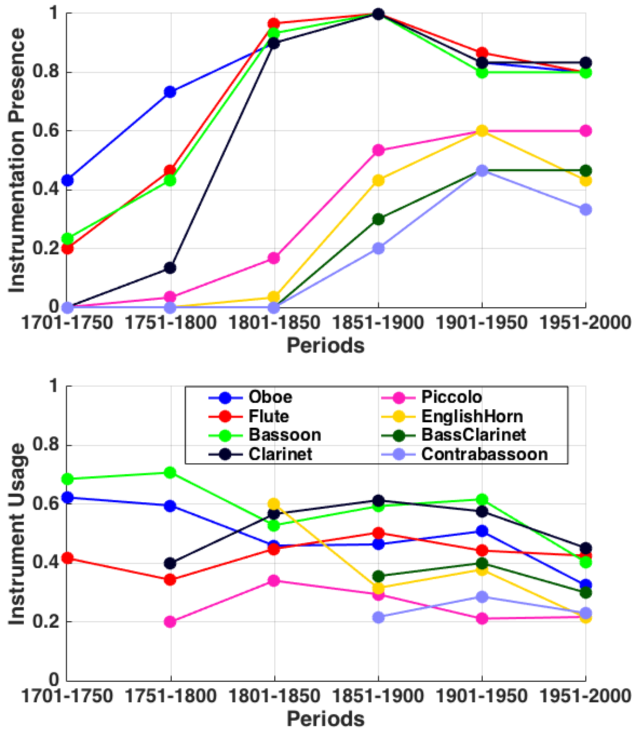


Figure 3. Instrumentation presence (top) and instrument usage (bottom) of woodwind instruments

The patterns of brass instruments are presented in Figure 4. It is interesting to see that the French horn shows the highest presence, which also tends to sound the most often. There could be various reasons behind this prominent presence and usage of the French horn. First, it became a part of an orchestra very early. Second, its timbre is multi-faceted, which makes the instrument quite versatile. For this reason many composers used the French horn in prominent solos, or to support other brass instruments, or to improve the blend of brass and woodwind instruments (Adler, 2002). Other brass instruments were increasingly more present until the early 20th century, after then their presence slightly decreased. This pattern is consistent with what we observed in the total number of parts in Figure 1.

Figure 5 presents the patterns of percussion and keyboard instruments. Perhaps the most striking is the decline of two instruments (continuo and harpsichord) in contrast to other instruments' increase in presence. Also noticeable is the higher presence of timpani above all the other instruments, although it sounded only about 30% of the duration of a musical work. In terms of usage patterns, we see that harpsichord, organ, and continuo in the 18th century were nearly omnipresent, probably because they all performed the role of continuo. It is not clear from the graph whether the piano preserved the tradition of a keyboard continuo after the decline of harpsichord and continuo parts, or it just happened to show a higher usage pattern in the 19th and early 20th centuries. Figure 5 also shows that even though more percussion instruments appeared in orchestration after the early 19th century, their use was reserved for occasional special effects.

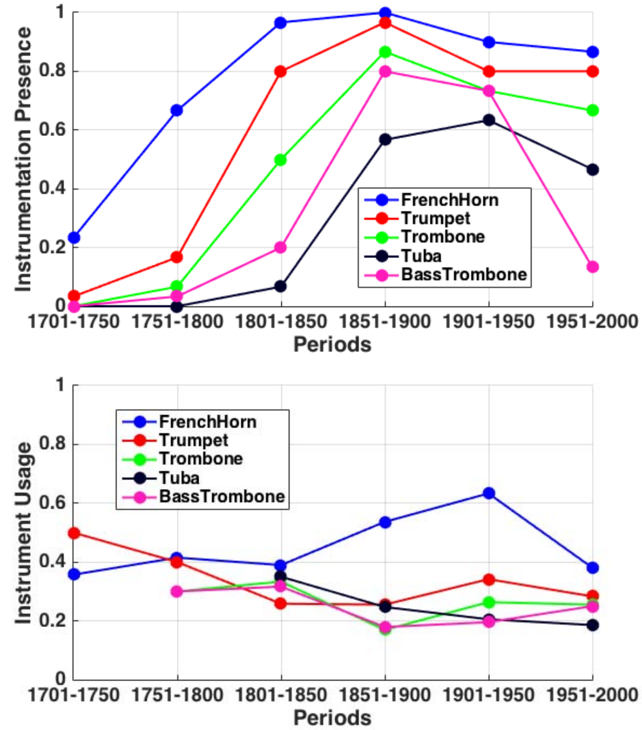


Figure 4. Instrumentation presence (top) and instrument usage (bottom) of brass instruments

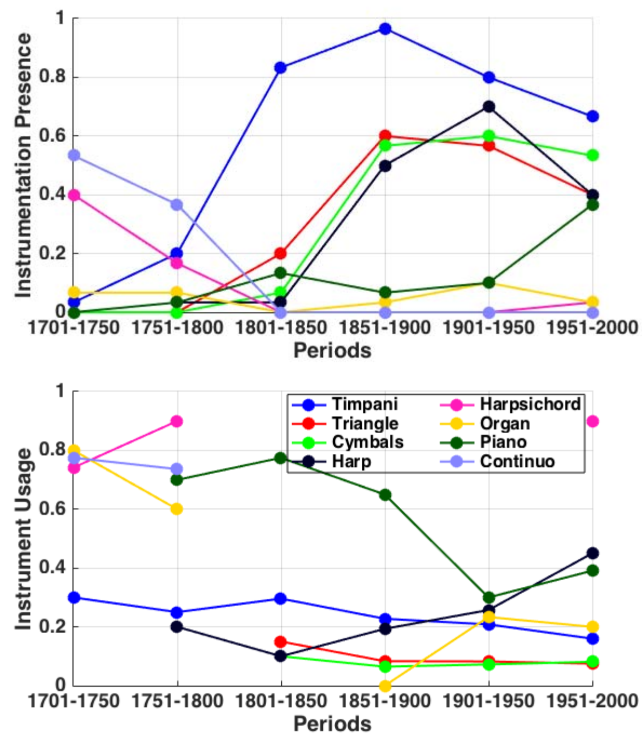


Figure 5. Instrumentation presence (top) and instrument usage (bottom) of percussion and keyboard instruments

We also examined the most common usage patterns found in our samples. Tables 2 and 3 list the results in descending order of occurrences. The standard string section combination (two violins, viola, cello, and contrabass) was the most common pattern over the 300-year period, although this exact

Table 2. Most common usage patterns in 300 years

Number of instances	Instrument combination pattern
67	Violin + Violin + Viola + Cello + Contrabass
27	Violin + Violin + Viola + Cello + Contrabass + Oboe + Oboe
20	Violin + Violin + Viola + Cello + Contrabass + Harpsichord
19	Violin + Violin + Viola
19	Violin + Violin
19	Violin + Violin + Viola + Cello
18	Violin + Violin + Viola + Continuo
17	Piano
14	Violin
12	Violin + Violin + Viola + Cello + Contrabass + Bassoon
10	Violin + Violin + Violin + Viola + Cello + Contrabass
10	Violin + Violin + Continuo
10	Cello + Contrabass
9	Violin + Violin + Violin
9	Violin + Violin + Viola + Cello + Contrabass + Piano
9	Violin + Violin + Viola + Cello + Harpsichord + Continuo

Table 3. Most common usage patterns per epoch

Epoch	Number of instances	Instrumentation pattern
1701 – 1750	19	Violin + Violin + Viola + Cello + Contrabass + Harpsichord
	15	Violin + Violin + Viola + Continuo
	14	Violin + Violin + Viola + Cello + Contrabass + Oboe + Oboe +Bassoon
1751 – 1800	27	Violin + Violin + Viola + Cello + Contrabass
	13	Violin + Violin + Viola + Cello + Contrabass + Oboe + Oboe + Bassoon
	9	Violin + Violin + Viola + Cello + Contrabass + Harpsichord + Continuo
1801 – 1850	10	Violin + Violin + Viola + Cello + Contrabass
	9	Piano
	7	Violin + Violin + Viola + Cello + Contrabass + Piano
1851 – 1900	4	Violin + Violin + Viola + Cello
	3	Violin
	3	Violin + Violin + Viola + Cello + Contrabass
1901 – 1950	17	Violin + Violin + Viola + Cello + Contrabass
	6	Violin + Viola + Cello + Contrabass
	5	Violin + Violin + Viola + Cello
1951 – 2000	5	Piano
	4	Violin + Violin + Violin + Violin + Viola + Viola + Cello + Contrabass
	3	Violin + Violin + Viola + Cello + Contrabass
	3	Cello
	3	Harp

pattern was not very popular in 1701−1750. Table 3 shows that the use of string instruments showed a slight decrease through the years, in contrast to wind instruments.

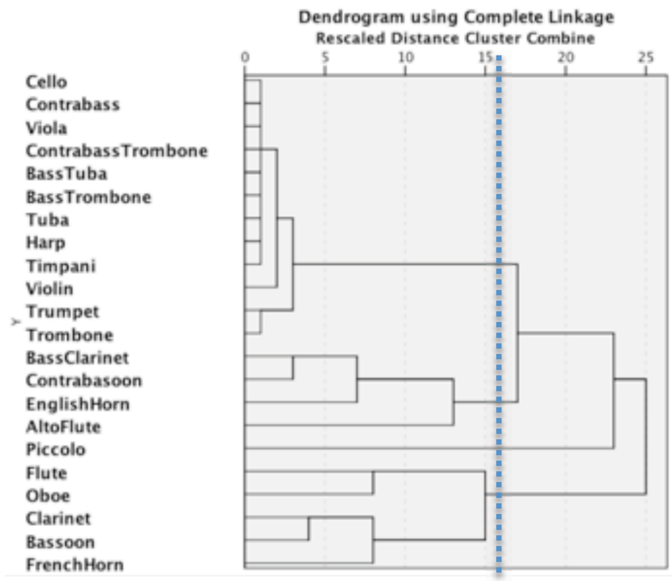


Figure 6. Hierarchical clustering of 22 orchestral instruments

Interested in how various instruments were combined with each other, we carried out a hierarchical clustering analysis on the usage patterns of every pair of 22 typical orchestral instruments (Violin, viola, cello, contrabass, piccolo, flute, alto flute, oboe, English horn, clarinet, bass clarinet, bassoon, contrabassoon, French horn, trumpet, trombone, bass trombone, contrabass trombone, tuba, bass tuba, harp, and timpani). Figure 6 shows the resulting dendrogram. At around the distance of 16, we observe four clusters: on the top is a big cluster containing all string instruments and all brass instruments except for the French horn, plus harp and timpani (cello, contrabass, viola, contrabass trombone, bass tuba, bass trombone, tuba, harp, timpani, violin, trumpet, and trombone). These instruments probably form the core of the orchestra. Next is a cluster of four woodwinds offering extended pitch and timbral color (bass clarinet, contrabassoon, English horn, and alto flute). Strangely, piccolo is in its own cluster for some unknown reason. The last cluster consists of five wind instruments (flute, oboe, clarinet, bassoon, and French horn) that showed higher presence than others (evident in Figures 2 and 3). If we take the liberty to include piccolo in the second cluster with other woodwinds, we have three clusters that reflect potentially different compositional functions. The first cluster might be called the *Core* cluster, where all string instruments seems to provide the core of an orchestral scene; the second cluster contains five woodwinds (bass clarinet, contrabassoon, English horn, alto flute, and piccolo) that offer extended pitch and timbre ranges, therefore dubbed the *Extended Woodwind* cluster; the third cluster might be deemed the *Standard Wind* cluster as it contains five wind instruments that are rather standard in orchestras (flute, oboe, clarinet, bassoon, and French horn). The second and third clusters probably reflect different purposes in instrumentation of woodwinds; the instruments in the third cluster tend to be used more consistently, complementing those in the core cluster, whereas those in the second cluster are more likely to be used for specific effects in pitch and timbre. Our clusters did not perfectly agree with Johnson's (2011) SPC model. However, like Johnson's model, our clusters nevertheless appear to point to different compositional functions.

IV. DISCUSSION

In this paper, we surveyed the instrumentation patterns in orchestral music composed between 1701 and 2000, concentrating on instrumentation presence and instrument usage. From our samples we could clearly see the constant increase in the total number of parts required to perform a piece of music through history up until the mid-20th century. In contrast to the increase of the number of parts in an orchestral work, the usage pattern of an instrument showed a decreasing trend in general. This decline in usage may simply reflect the increased palette of instruments from which a composer might choose; that is, when the composer can choose from many instruments, any given instrument need not play constantly.

In terms of instrumentation presence, string instruments appear to show the highest presence. Strings have been a core of an orchestra since the 18th century and therefore present in nearly every piece of orchestral music (Weaver, 2006). The wind instruments in current orchestras seem to belong to three groups depending on the compositional purposes they serve—the *Standard Wind* group (flute, oboe, clarinet, bassoon, and French horn), the *Core* (brass) group (trumpet, trombone, bass trombone, contrabass trombone, tuba, bass tuba), and the *Extended Woodwind* group (bass clarinet, contrabassoon, English horn, alto flute, and piccolo). The instruments in the Standard Wind group have been available for a longer time and they tend to sound more often than others. These instruments offer a set of stable timbres to the orchestra, which steadily increased with the inclusion of instruments that provide more diversity in timbre and pitch (especially those in the extended woodwind group). The Core brass instruments add more weight to the skeleton of an orchestral image painted by the string instruments. Harp and timpani, which belong to the same cluster, also appear to serve the same purpose of adding weight. Our three clusters did not entirely replicate Johnson's SPC model, possibly because our data included samples from a longer period than the Romantic era, whereas Johnson focused exclusively on the Romantic period.

V. CONCLUSION

To our knowledge, this is the first empirical longitudinal study of orchestration. The general patterns we found mostly agree with what have been qualitatively observed by music scholars (Harnoncourt, 1988; Stauffer, 2006; Weaver, 2006). Many of our analyses were performed on the entire set of samples for the purpose of statistical power, but it might have led to glossing over possible changes observable per epoch. Therefore our findings need to be interpreted carefully. Although this study required extensive labor to sample and code 1800 sonorities from 180 works, a larger collection of data is necessary in order to survey detailed longitudinal trends. Further analysis of presence and usage patterns in relation to pitch, tempo, and dynamics are expected to follow, as well as examination of occurrences of specific patterns mentioned in orchestral treatises.

REFERENCES

Adler, S. (2002). The Study of Orchestration. New York: W.W. Norton.

Berlioz, L.H. (1855). Berlioz's orchestration treatise: A translation and commentary (Cambridge musical texts and monographs) (H. Macdonald, Ed.). Cambridge, England: Cambridge University Press. (Republished in 2002).

Dubal, D. (2001). The Essential Canon of Classical Music. New York: North Point Press.

Harnoncourt, N. (1988). Baroque music today: Music as speech: Ways to a new understanding of music (R.G. Pauly, Trans.). Portland, Oregon: Amadeus Press.

Johnson, R. (2011). The Standard, Power, and Color model of instrument combination in Romantic-era symphonic works. Empirical Musicology Review, Vol. 6, No. 1, pp. 2-19.

Jolly, J. (2010). The Gramophone Classical Music Guide. Haymarket Consumer Media.

Koechlin, C. (1954-9). Traité de l'orchestration: en quatre volumes. Paris: M. Eschig.

Piston, W. (1955). Orchestration. New York: W.W. Norton.

Rimsky-Korsakov, N. (1964). Principles of Orchestration. Edited by Maximilian Steinberg. English translation by Edward Agate New York: Dover Publications. (Original work published 1912)

Stauffer, G.B. (2006). The modern orchestra: A creation of the late eighteenth century. In J. Peyser (ed.), The Orchestra (pp. 41-72). Milwaukee, WI: Hal Leonard.

Weaver, R.L. (2006). The Consolidation of the main elements of the orchestra: 1470-1768. In J. Peyser (ed.), The Orchestra (pp. 7-40). Milwaukee, WI: Hal Leonard.

Widor, C.-M. (1904/1906) Technique de l'orchestre moderne. Paris, 1904. Trans. by Edward Suddard as: The technique of the modern orchestra: A manual of practical instrumentation. London, J. Williams, 1906.

Reduced cultural sensitivity to perceived musical tension in congenital amusia

Cunmei Jiang[1], Fang Liu[2], and Patrick C.M. Wong[3]
[1]Music College, Shanghai Normal University, Shanghai, China
[2]Department of Language and Linguistics, University of Essex, Colchester, Essex, UK
[3]Department of Linguistics and Modern Languages and Brain and Mind Institute, The Chinese University of Hong Kong, Hong Kong SAR, China

Everyday music listeners demonstrate reduced sensitivity to tension in music from an unfamiliar culture relative to their own culture. This in-culture bias indicates not only perceptual learning, but also an enhanced emotional interaction with what has been implicitly learned.

The aim of the present study is to investigate whether amusic individuals demonstrate such an in-culture bias as to assess whether implicit learning and emotional responses to what has been implicitly learned constitutes a deficit domain in amusia. The literature reporting on core deficit domains in amusia typically focuses on musical expectancy including evaluating melodic and harmonic structures. Yet how failing at musical expectancy may lead to reduced sensitivity to more global musical phenomena such as culture and tension is unknown. We seek to examine whether sensitivity to such global musical phenomena is impaired in amusia.

Twenty-six amusics and 26 matched controls listened to 32 short Western and Indian melodies, played either on the piano or sitar, with or without syntactic violations (two in-key and two out-of-key variants), and were asked to rate the tension induced by each melody using a Continuous Response Digital Interface dial.

A four-way ANOVA on tension ratings revealed a significant interaction between group, sound, and music type. Although amusics rated tension differently depending on piano or sitar sounds, they were unable to distinguish the difference in tension between Western and Indian music. In contrast, controls gave different ratings for tension in Western versus Indian music for both piano and sitar sounds.

These findings indicate that amusics have reduced cultural sensitivity to perceived tension in music. Since musical tension rating taps into emotional responses, this suggests that amusics may also have impaired affective responses to music.

A Comparison of Statistical and Empirical Representations of Cultural Distance

Steven J. Morrison[1], Steven M. Demorest[2], and Marcus Pearce[3]
[1]Laboratory for Music Cognition, Culture & Learning, University of Washington, USA
[2]Bienen School of Music, Northwestern University, USA
[3]School of Electronic Engineering and Computer Science, Queen Mary University of London, UK

Individuals, regardless of music training, demonstrate better memory for music that is culturally familiar than music that is culturally unfamiliar. Among Western listeners, memory advantage for own-culture music persists even when music examples are reduced to isochronous pitch sequences of uniform timbre, suggesting that melodic expectancy in the form of the pitch-to-pitch intervallic content may be a useful measure of enculturation. We have hypothesized that the cognitive distance between the musics of two cultures may be represented by differences in the statistical content of melodies from those cultures.

In the first part of this study we tested this hypothesis by examining a corpus of Western and of Turkish melodies to determine if culture-specific typicality of interval content, melodic contour, rhythm content, or some combination of these parameters showed a relationship to cross-cultural memory performance. Interval content demonstrated the strongest correlation with preexisting memory performance data.

In the second part of this study currently underway we further test this finding. Using melodies that demonstrated the strongest typicality within their respective corpora, adult listeners will complete an open sorting task in which they will organize 24 Western and Turkish melodies into two categories of their own devising. This will be followed by a recall test in which participants will identify whether brief music excerpts were or were not taken from the sorted melodies. We predict that (1) sorting differentiation will best reflect culture-specific measures of pitch content, (2) memory will be more accurate for melodies derived from the culturally familiar corpus, and (3) memory performance will demonstrate the strongest relationship with own-culture pitch content typicality.

A cross cultural study of emotional response to North Indian Classical Music

Avantika Mathur[1], Vishal Midya[2], Nandini C. Singh[1]
[1]Language, Literacy and Music lab, Cognitive Neuroscience, National Brain Research Centre, Manesar, India
[2]Indian Statistical Institute, Kolkata, India

Emotion occupies centre stage amongst the numerous expressive qualities of music. Recently, it was shown that *ragas* which are specific combinations of tonic intervals are capable of eliciting distinct emotions. It was also shown that *alaap* and *gat*, the two modes of presentation of a *raga* that differ in rhythmic regularity, differ in the level *of emotion* arousal. While tonality (estimated in terms of minor to major ratio) determined the emotion experienced for a *raga*, rhythmic regularity and tempo modulated level of arousal. The primary objective of this study was to investigate cultural similarities and differences in experienced emotion across enculturated and non-enculturated populations in such *raga* stimuli. 143 enculturated and 112 non-enculturated participants provided responses of 'experienced emotion' to 12 Hindustani classical *ragas* presented in *alaap* and *gat* on a Likert scale of 0-4. Analysis of responses revealed similar emotion experienced by the two groups, while factor analysis revealed differences in the way emotion labels were grouped together. 'Longing' was grouped with 'negative' factor (i.e. sad, tensed and angry emotions) by enculturated participants whereas it was grouped with 'positive' factor (i.e. happy, romantic, calm, and devotional) by the non-enculturated group. To investigate level of emotion arousal, ratings of highest experienced emotion were correlated with tonality for both *alaap* and *gat*. Results revealed significant correlations between tonality and emotion rating for both *alaap* and *gat* in enculturated participants, while non-enculturated participants showed a significant correlation only during *gat*. While our results suggest universality in emotions experienced by culturally distinct populations, the differences in level of emotion arousal suggest that rhythmic cues may be used similarly but tonality cues may modulate emotion response differentially by the two populations.

Relating Musical Features and Social Function: Contrast Pattern Mining of Native American Music

Daniel Shanahan[1], Kerstin Neubarth[2], and Darrell Conklin[3,4]
[1]School of Music, Louisiana State University, USA
[2]Canterbury Christ Church University, UK
[3]Dept of Computer Science & Artificial Intelligence, University of the Basque Country UPV/EHU, Spain
[4]IKERBASQUE, Basque Foundation for Science, Spain

One of the most striking benefits of the use of large corpora in music research has been the ability to examine relationships and commonalities between cultures, which have associated linguistic, emotional, and social universals underpinning musical functions. The current study takes an alternative approach, examining contrasts rather than commonalities within Native American music, while also discussing how data mining can allow for an understanding of the interaction between specific social functions and musical features.

A quantitative approach to Native American music predates computational approaches. Using a corpus of 334 songs, Gundlach (1932) examined the differences in music feature distributions related to conveyed emotions (as inferred by social function). Here we further investigate Gundlach's hypotheses using the computational methods of contrast pattern mining (Neubarth and Conklin, 2016). Data mining is applied to the Densmore collection of over 2000 Native American folk songs (Shanahan and Shanahan, 2014). Densmore associated her transcriptions with social functions, such as "game songs" and "hunting songs", which serve as a foundation for our own ontology of more general "superfunctions". Contrast pattern mining is used in our study to discover global feature patterns and sequential patterns which distinguish between social superfunctions.

In this study, we build upon Densmore's own analyses and Gundlach's early comparisons in Native American music to further explore the relationships between musical features and social function. For example, results reveal certain sequential patterns of melodic intervals and contour, as well as melodic interval global features, that are significantly overrepresented in love songs, game songs, and healing songs. These results confirm and extend earlier observations on Native American Music.

Effects of Repetitive Vocalization on Temporal Perception in a Group Setting

Jessica Pinkham,*1 Zachary Wallmark,*2

*SMU Meadows School of the Arts, Dallas, TX, USA

1jpinkham@smu.edu, 2zwallmark@smu.edu

I. BACKGROUND

Spiritual musicking is ubiquitous across human cultures (Ellingson, 1987). As evidenced in practices as diverse as Balinese bebutan trancing, Krishna kirtan singing, and Vedic mantra chanting, spiritual musicking has the capacity to modulate arousal levels and distort temporal perception (Becker, 2004; Black Brown, 2014). Though the cultural value and phenomenological efficacy of music in these and many other cultural practices is widely acknowledged by ethnomusicologists (Rouget, 1985), the perceptual correlates of spiritual musicking are poorly understood. In this study, we examined one shared trait among the practices in question: sustained, synchronized vocalization of a small number of syllables on a single pitch, as exemplified by mantra-style chant.

II. AIMS

We aimed to explore the effects of sustained, synchronized vocalization on temporal perception, in comparison to silence, listening, and speaking tasks. Accuracy of time estimates was used as an indicator of participants' retrospective experience of "clock time." The difference between subjects' estimates and the actual "clock time" that elapsed during each task indicated relative levels of temporal distortion experienced by the subjects. We hypothesized that the mantra-style vocalization (i.e., *chanting*) would produce the greatest variability of time estimations.

III. METHOD

Participants were two groups of students at SMU ($N = 58$), 28 male, 30 female; 35 music majors, 23 non-major; age 18–30 ($M = 19.55$, $SD = 2.09$); with 0–16 years of musical training ($M = 8.49$, $SD = 5.44$). The study consisted of four conditions: (1) sitting in silence as a group, (2) listening to a recording of a group chanting neutral syllables "shaun" (ʃa–ɔ–un), (3) speaking the neutral syllables "shaun" repetitively as a group, and (4) chanting "shaun" repetitively as a group. For each condition, participants were asked to estimate as accurately as possible in seconds how much time had elapsed during the performance of the task.

Conditions were presented in randomized order between the groups of students, and each task was randomly assigned a duration of 45s, 60s, 75s, or 90s. To compare data from the same tasks across different time intervals, we defined accuracy of guesses as the ± standardized differences (in percentage) of participants' time estimations from the actual time. Three outliers were removed due to a 100% or greater standardized difference from the actual time.

IV. RESULTS

A one-way repeated measures ANOVA revealed a significant effect of the task on temporal judgment, $F(3, 171) = 8.29$, $p < .0001$. As shown in Fig. 1, subjects consistently underestimated the duration of time that had elapsed: standardized mean difference (*SMD*) of time estimates for all four tasks was less than the actual time (silence [$SMD = -.28$], listening [$SMD = -.31$], speaking [$SMD = -.23$], and chanting [$SMD = -.15$]). However, while chanting produced the most "accurate" temporal evaluation according to group means, it exhibited the highest standard deviation by a substantial margin. Variability increased linearly from silence to chanting tasks (silence [$SD = .18$], listening [$SD = .21$], speaking [$SD = .22$], and chanting [$SD = .32$]). The high variability of time estimations during the chanting task provides tentative support for our hypothesis that synchronized repetitive vocalization alters temporal perception to a greater degree than listening or speaking.

Music majors were more accurate at time estimation than non-majors during all tasks (music majors $SMD = -.17$, non-majors $SMD = -.33$), and the chanting mean for music majors was the most accurate of the whole dataset ($SMD = -.01$). This finding held when assessing the relationship between time estimation and years of musical training as well: musical training is weakly though significantly correlated with estimation accuracy during silence ($r = .31$, $p = .02$) and chanting ($r = .36$, $p < .01$).

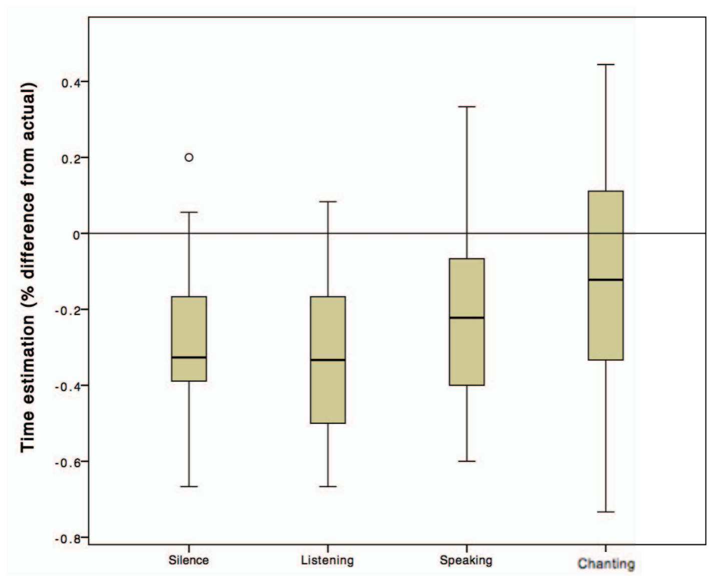


Figure 1: Time estimation ($N = 58$) in four conditions

V. CONCLUSION

This study provides preliminary insight into the ubiquity and efficacy of chant practices. Sustained group vocalizations may more effectively lead to a distortion of time perception compared to merely speaking or listening to the same syllables.

REFERENCES

Becker, J. (2004). *Deep Listeners: Music, Emotion, and Trancing.* Bloomington, IN: Indiana University Press.

Black Brown, S. (2014). Krishna, Christians, and colors: The socially binding influence of Kirtan singing at a Utah Hare Krishna festival. *Ethnomusicology, 58*(3). 455-474.

Ellingson, T. (1987) "Music and Religion." In *Encyclopedia of Religion,* Vol. 9, ed. M. Eliade. London: Macmillan.

Rouget, G. (1985) *Music and Trance*, trans. B. Biebuyck. Chicago: University of Chicago Press.

Empirical Musicology, 10 Years On: New Ways to publish and the empirical paradigm in music research

Daniel Shanahan[1]
[1]School of Music, Louisiana State University, Baton Rouge, LA, USA

2016 represents the 10th anniversary for the open-access journal *Empirical Musicology Review (http://emusicology.org/),* which has followed the open peer-review model of *Behavioral and Brain Sciences,* in which articles are published alongside commentaries by peers, in an attempt to open the lines of discourse in the field. Since the journal began, however, the academic publishing world has seen seismic shifts in attitudes toward access, peer-review, and the publishing model in general. At the same time, empirical musicology as a research field has developed substantially over the past ten years along with developments in neighbouring areas such as music information retrieval, music and neurosciences, embodied music cognition and digital musicology. These neighbouring research areas all subscribe to the empirical research paradigm and the idea that musicological and scientific insights need to be based on empirical evidence. However, while they share the empirical approach to music research, their individual research perspectives and their ways of presenting themselves as current areas of music research differ markedly. Given this diversity of empirical approaches to music research, can and should empirical musicology provide a conceptual bracket that facilitates exchange of research ideas and empirical findings between these areas?

This symposium will address several short-topic questions related to the recent changes in scientific publishing, as well as on the role of empirical musicology as a paradigm for modern music research. Topics will be presented by several contributors in short position papers (5-7 minutes), followed by a roundtable discussion. Specifically, topics will focus on the role of open-access in music research, what questions empirical musicology should seek to answer, what tools and methods are currently lacking, the benefits and disadvantages of open-peer review and a broader focus on the past, present, and future of empirical musicology.

Although this deviates from a typical symposium consisting of longer talks, we are hopeful that the representation of scholars with varying interests and from many different countries will provide an equally rich discussion.

The History of Empirical Musicology Review

David Huron[1]
[1]School of Music, Ohio State University, Columbus, OH, USA

What questions should we in empirical musicology be seeking to answer?

Marc Leman[1]
[1]IPEM, Department of Musicology, University of Ghent, Belgium

What is the point of empirical musicology if we already have music theory, music informatics and music psychology?

Justin London[1]
[1]Department of Music, Carleton College, Northfield, MN, USA

What tools and methods are we lacking?
Reinhard Kopiez[1]
[1]University of Music, Drama and Media, Hanover, Germany

How can music theory, music psychology and music informatics strengthen each other as disciplines?
Anja Volk[1]
[1]Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands

Where to draw the line between 'empirical' and 'non-empirical' in musicology
Peter Keller[1]
[1]The MARCS Institute, University of Western Sydney, Australia

Can new publishing models benefit researchers who read and respond constructively to work by others?
Alan Marsden[1]
[1]Department of Music, Lancaster University, Lancaster, UK

Does a great empirical music paper need to be published in a high-impact factor journal?
Annabel Cohen[1]
[1]Department of Psychology, University of Prince Edward Island, Charlottetown, PE Canada

How and why does empirical data make music research more impactful
Alexandra Lamont[1]
[1]School of Psychology, Keele University, Keele, UK

Where will empirical musicology venture over the next 10 years and what does it take to get there?
Elizabeth Hellmuth Margulis[1]
[1]School of Music, University of Arkansas, Fayetteville, AR, USA

Will the next generation of musicologists be all empirical musicologists?
Eric Clarke[1]
[1]Faculty of Music, University of Oxford, Oxford, UK

Empirical Musicology Futures
Nicola Dibben[1]
[1]Department of Music, The University of Sheffield, Sheffield, UK

The future of the specialised music research journal in the age of digital publishing
Daniel Müllensiefen[1]
[1]Department of Psychology, Goldsmiths, University of London, London, UK

Perspectives on the History of Music Cognition

Amy B. Graziano[1] and Julene K. Johnson[2]
[1]Hall-Musco Conservatory of Music, Chapman University, Orange, CA, USA
[2]Institute for Health & Aging, University of California, San Francisco, CA, USA

The field of music cognition has developed for approximately 150 years. However, extended discussion of its origins and growth in a historical sense has emerged only recently. Investigating how and why music became one of the domains of brain research in the twentieth and twenty-first centuries can lead to a deeper understanding of psychology and cognitive science today. Exploring the roots of music research takes us back to the nineteenth century. Most music psychologists agree there was an increasing interest in music and brain in the late nineteenth century, particularly in the work of Hermann von Helmholtz and the beginnings of psychoacoustics. The late nineteenth century also saw the emergence of musicology, psychology and neurology as new fields of study. Literature in these areas shows that Helmholtz was not alone, but that in fact, researchers in all of these disciplines were interested in music and the brain. By the end of the nineteenth century, the study of music and brain had formed a separate identity within the fields of neurology, psychology and musicology. Many issues that were explored in the late nineteenth and early twentieth centuries have become part of twentieth-first century investigations in music cognition and in neuroscience: localization of music function, music impairments, perceptual and cognitive processing, the relationship between music and emotion. During the early and mid-twentieth century, however, the field of music cognition did not continue to grow as rapidly as during the late-nineteenth century. The first talk will focus on the contributions of neurologists to the study of music cognition in the late nineteenth century. The second talk will explore the overlapping interest of late nineteenth-century neurologists, psychologists, and musicologists who shared an interest in music cognition. The final talk will focus on the impact of American Behaviorism on the field of music psychology during the early and mid-twentieth century.

Nineteenth-Century Roots of Music Cognition: Influence of Neurology

Julene K. Johnson[1] and Amy B. Graziano[2]
[1]Institute for Health & Aging, University of California, San Francisco, CA, USA
[2]Hall-Musco Conservatory of Music, Chapman University, Orange, CA, USA

The historical roots of the current interest in studying music and brain can be traced back to the second half of the nineteenth century when neurology and psychology were gaining momentum as respected fields. During this time, music was used to examine and better understand higher cognitive functioning in persons with brain damage, particularly those with aphasia (i.e., difficulty speaking or understanding language). Early neurologists were fascinated by the paradox that some patients who were unable to speak could sing. This seemingly simple observation inspired decades of clinical inquiry about music and brain, and these early neurologists developed theories about how music was processed by the brain. The purpose of this talk is to summarize observations of music and brain through the eyes of three important nineteenth-century neurologists: August Knoblauch, Hermann Oppenheim, and Jean-Martin Charcot. We used historical methods to review archival material. The presentation will discuss similarities and differences among these three figures. Music played an important role in early thinking about cognition and brain and helped shape the emerging field of neurology. After various music abilities were investigated over time and with different patients, it became apparent that music involved a complex cognitive system and brain functions.

Nineteenth-Century Roots of Music Cognition: Interdisciplinary Perspectives

Amy B. Graziano[1] and Julene K. Johnson[2]
[1]Hall-Musco Conservatory of Music, Chapman University, Orange, CA, USA
[2]Institute for Health & Aging, University of California, San Francisco, CA, USA

As part of a new interest in how the brain processes music, prominent nineteenth-century neurologists began conducting investigations of music cognition in patients with neurological impairments. After careful observations of many patients, these German, French and English scientists identified a new neurological syndrome referred to as "amusia"—a systematic breakdown of musical abilities as a result of brain damage. This led to discussion of localization of music function within the brain. At the same time, investigations by psychologists and musicologists discussed similar concepts. However, research by psychologists and musicologists followed a different path from the neurologists, one focused on multiple levels of sensory processing and mental representation. The purpose of this talk is to summarize the work of nineteenth-century psychologists and musicologists who were interested in music and brain, in particular Carl Stumpf and Richard Wallaschek, and to discuss parallels with the work of contemporary neurologists (Knoblauch and Oppenheim, discussed in the first talk, and English neurologist John Hughlings Jackson). We used historical methods to review archival material that reveals music played a greater role in nineteenth-century scientific thought within psychology and neurology than has been previously acknowledged.

The Impact of Behaviorism on the History of Music Psychology

Annabel J. Cohen[1]
[1]Department of Psychology, University of Prince Edward Island, Charlottetown, Canada

Today's favorable climate for music psychology contrasts with the climate found within early and mid twentieth-century American Behaviorism. Behaviorism was both a negative reaction to earlier decades that had founded a seemingly unproductive human introspective psychology and a positive reaction to the success of Pavlov's learning laws derived through animal research. Between 1913 and 1975 (the cusp of the Cognitive Revolution), Behaviorism promoted the study of "objectively observable behavior" and demoted studies of mind and imagery. New laboratories for experiments with rats and pigeons created an atmosphere unconducive to music research. A few researchers such as Carl Seashore soldiered on in the music realm. But PsycInfo reveals a dip in the number of papers referring to "music" (as compared to "psychology" and "behavior", for example) during this Behaviorist period. Several milestones mark a return to the legitimacy of music psychology. In 1969, Diana Deutsch published an article in *Psychological Review* entitled "Music Recognition, applying a hierarchical physiological model of feature detectors to account for melodic transposition". There followed under the editorship of George Mandler several more articles on music themes. In 1978, Roger Shepard organized a symposium on the cognitive foundations of music pitch at the annual meeting of the Western Psychological Association. It took place in this very hotel, introducing, to a standing-room-only crowd, Carol Krumhansl and the tonal hierarchy along with several other (then young) researchers active to this day. Thus began the post-Behaviorism legitimacy of music cognition research. While this recognition eventually became more widespread, specialty journals sprang up. In spite of the downside to Behaviorism, music psychology benefited indirectly from Behaviorism's rigor and developments in statistics. The historical context highlights the privilege it is to study music psychology in the current cognitive/neuroscience zeitgeist.

Musical structure as a hierarchical retrieval organization: Serial position effects in memory for performed music

Jane Ginsborg[1], Roger Chaffin[2], Alexander Demos[3] and James Dixon[2]
[1]Royal Northern College of Music, Manchester, UK
[2]Psychology Department, University of Connecticut, Storrs, CT, USA
[3]Psychology Department, University of Illinois, Chicago, IL, USA

When musicians first learn pieces of music, serial cuing activates memory for what follows, as each passage played or sung is linked to the next. During memorization, serial memory is overlaid with different types of performance cue (PCs) enabling musicians to recall a specific passage, independent of musical context, simply by thinking of it, e.g., “coda”. The locations of retrieval cues can be identified by the presence of serial position effects in free recall. We looked for such effects in the written recall of an experienced singer (the first author) who wrote out the score of a song memorized for public performance six times over a period of five years. The singer recorded all of her practice. After the performance she marked the locations of phrases and sections in the score, and the locations of the PCs that she had used. Previously, we have reported how serial position effects for PCs were different for different types of PC. Here, we use logistic and growth-curve mixed models to examine how serial position effects for PCs were affected by practice, showing how recall changed over time for better and worse. Initially, recall was perfect. Fourteen months after the performance, there was a normal primacy effect at Expressive PCs marking musical feelings, and an inverse primacy effect at Word PCs marking nuances of stress and pronunciation. By the time of the last recall, five years after performance, these effects had reversed. The normal effect for Expressive PCs became inverse, and vice versa for Word PCs; and there was also an inverse effect for Basic PCs marking details of technique such as breathing. PCs acted as retrieval cues, providing content addressable access to an associatively chained memory and countering the effect of weakening links between passages. The direction of primacy effects depended on the salience of each type of PC, which changed over time with the singer’s interest in the piece, initially as performer and later as researcher.

How accurate is implicit memory for the key of unfamiliar melodies?

E. Glenn Schellenberg[1] and Michael W. Weiss[1]
[1]Department of Psychology, University of Toronto Mississauga, Mississauga, ON, Canada

Musically untrained listeners remember the key of previously unheard melodies, but the key manipulations in earlier research were large (6 semitones). Here we sought to document the accuracy of implicit memory for key. Our 24 stimulus melodies were taken from the vocal lines of unknown Broadway songs and presented in piano timbre. Each was saved in five different keys. The standard (lowest) key had a median pitch of G4. Other versions had a median pitch 1, 2, 3, and 6 semitones (ST) higher than the standard. Musically untrained participants were tested in one of four conditions (1, 2, 3, and 6 ST) based on the magnitude of the key change. Before the experiment began, listeners learned that key is irrelevant to a melody's identity. During an exposure phase, listeners heard 12 of 24 melodies (selected randomly): 6 in the standard key, 6 in the higher key. After a 10-min break, they heard all 24 melodies. The 12 new melodies were divided equally among the standard and higher keys. Six of the 12 old melodies were presented in the original key. The other six were transposed from the standard to higher key or vice versa. Listeners rated how confident they were that they had heard the melody in the exposure phase. Across conditions, ratings were much higher for old than new melodies, indicating that the melodies were remembered well. Memory for old melodies was also superior when they were re-presented in the original key, indicating memory for key. This original-key advantage was evident in the 2, 3 and 6 ST conditions. It disappeared, however, in the 1 ST condition. Overall recognition (ignoring the key change) was also better in the 1 and 2 ST conditions than in the 3 and 6 ST conditions. The findings indicate that (1) implicit memory for the key of previously unfamiliar melodies is remarkably accurate, and (2) memory for a corpus of novel melodies is superior when the overall pitch range is restricted.

Effects of performance cues on expressive timing

Tânia Lisboa[1], Alexander P. Demos[2], Kristen Begosh[3], and Roger Chaffin[4]
[1]Centre for Performance Science, Royal College of Music, London, UK
[2]Department Psychology, University of Illinois at Chicago, IL, USA
[3]Department Psychology, University of Delaware, Newark DE, USA
[4]Department Psychology, University of Connecticut, Storrs, CT, USA

Background
Performance cues (PCs) are the landmarks of a piece of music that a performer attends to during performance. Experienced performers strategically select aspects of the music to attend to during performance in order to achieve their musical and technical goals; these are their PCs. While most aspects of performance become automatic with practice, PCs provide musicians with a means of consciously controlling the otherwise automatic motor sequences required in performance.

Aim & Method
Support for this claim comes from a longitudinal case study in which an experienced cellist (the first author) recorded her practice as she learned the Prelude from J.S. Bach's Suite No. 6 for solo cello over a two-year period. In addition to recording 35^{+} hours of practice, she also recorded 21 practice, 7 live, and 8 "lab" performances in which the cellist played with normal or minimal expression. We have previously reported that playing during these performances was consistently slower at expressive and interpretive PCs. In the present study, we examined the effects of PCs on the bar-to-bar tempo of the eight lab performances.

Results & Conclusions
Slowing occurred at expressive and interpretive PCs and these effects were larger in normal than in non-expressive lab performances, suggesting that they were not due to problems with memory retrieval. Also, tempo varied across performances more at expressive and interpretive PCs than at other locations in the piece; and for interpretive PCs this effect was also larger in normal than in non-expressive performances. These effects suggest that PCs allow musicians to introduce spontaneity and variety to their performances by providing the means to modulate the highly practiced, automatic motor sequences required to perform reliably in public.

A non-musician with severe Alzheimer's Dementia learns a new song

Amee Baird[1], Heidi Umbach[1] & William Thompson[1]
[1]Department of Psychology and ARC Centre for Excellence in Cognition and its Disorders, Macquarie University, Sydney, New South Wales, Australia

The hallmark symptom of Alzheimer's Dementia (AD) is impaired memory, but memory for familiar music can be relatively spared. We investigated whether a non-musician with severe AD could learn a new song and explored the mechanisms underlying this ability. NC, a 91 year old woman with severe AD (diagnosed 10 years ago) who loved singing but had no formal music training, was taught an unfamiliar Norwegian tune (sung to 'la') over 3 sessions. We examined (1) immediate learning and recall, (2) 24 hour recall and re-learning and (3) 2 week delayed recall. She completed standardised assessments of cognition, pitch and rhythm perception, famous melody recognition and a music engagement questionnaire (MEQ, informant version), in addition to a purposefully developed two word recall task (words presented as song lyrics, a proverb or as a word stem completion task). She showed impaired functioning across all cognitive domains, especially verbal recall (0/3 words after 3 minute delay). In contrast, her pitch and rhythm perception was intact (errorless performance of 6 items each) and she could identify familiar music. Her score on the 'response' (responsiveness to music) scale of the MEQ was in the superior range relative to healthy elderly participants. On the experimental tasks she recalled 0/2 words that were presented in song lyrics or proverbs, but correctly completed both word stems, suggesting intact implicit memory function. She could sing along to the learnt song on immediate and delayed recall (24 hours and 2 weeks later, both video recorded), and with prompting could sing it alone. On objective ratings of pitch and rhythm accuracy during song recall she showed better recall of the rhythm. This is the first study to demonstrate preserved ability to learn a new song in a non-musician with severe AD. Underlying mechanisms may include intact implicit memory, superior responsiveness to music and increased emotional arousal during song compared with word learning.

Vocal melodies are remembered better than piano melodies but sine-wave simulations are not

Michael W. Weiss[1], Sandra E. Trehub[1], and E. Glenn Schellenberg[1]
[1]Department of Psychology, University of Toronto Mississauga, Mississauga, Ontario, Canada

Listeners remember vocal melodies (*la la*) better than instrumental melodies (Weiss et al., 2012, 2015a, b), but it is unclear which acoustic features are responsible for the memory advantage. To date, the instrumental timbres tested (piano, banjo, marimba) have had fixed pitches and percussive amplitude envelopes, in contrast to the voice, which incorporates dynamic variation in pitch and amplitude, rich spectral information, and indexical cues about the singer. In the present study, we asked whether artificial stimuli with voice-like pitch and amplitude variation could elicit a memory advantage like that observed with natural vocal stimuli. Vocal renditions from Weiss et al. (2012) were used as source materials for generating "sine-wave" renditions in PRAAT. Specifically, continuous pitch (fundamental frequency) and amplitude information were extracted from the natural vocal performances and applied to a continuous tone of equivalent duration. The resulting melodies sounded more like dynamic theremin performances than like human singing. In the familiarization phase, listeners heard 16 melodies (half natural or sine-wave voices, half piano) and rated their liking of each melody. Half of the listeners received naturally sung melodies and half received sine-wave versions ($n = 30$ per group). In the test phase, listeners heard the same melodies and 16 timbre-matched foils, and they judged whether each was old or new on a 7-point scale (definitely new to definitely old). Listeners exhibited better memory for natural vocal melodies than for piano melodies ($p = .012$), as in previous research, but they performed no differently on sine-wave melodies and piano melodies ($p = .631$). These findings suggest that dynamic pitch and amplitude information from singing, in the absence of natural vocal timbre, is insufficient to generate enhanced processing. Instead, such processing advantages may depend on the full suite of acoustic features that is unique to human voices.

Perception of Leitmotives in Richard Wagner's *Der Ring des Nibelungen*

David John Baker[1] and Daniel Müllensiefen[2]
[1]Music Department, Louisiana State University, USA
[2]Psychology Department, Goldsmiths, University of London, UK

Listeners naive to a musical genre are thought to process structures on the musical surface differently than experts familiar with the genre. However, only a small amount research has been done exploring claims of differences in musical processing abilities associated with expertise beyond musical training. In this paper, we investigate the perception of leitmotives in Wagner's *Der Ring des Nibelungen* in an two experiments requiring listeners to recall musical material based on a short excerpt. We hypothesized that both individual differences as well as musical parameters of the leitmotives would contribute to the memory performance of the listeners. In the experiment participants (N=100 and N=30) listened to a ten minute excerpt from the opera *Siegfried* and completed a memory test to indicate which leitmotives they remembered hearing in the excerpt. The independent variables measured were musical training, German speaking ability, measures of Wagner affinity and knowledge , and the acoustic similarity determined chroma distance. As dependent variables participants provided ratings of implicit and explicit memory as well as affective ratings for each leitmotive in the memory test. A structural equation model of the data indicates that Wagnerism, a latent trait combining affinity to and knowledge of Wagner's music is almost twice as strong as a predictor for participants' leitmotive memory than musical training. Regression analysis of the musical parameters indicated that acoustic chroma was a significant predictor for the individual difficulty of leitmotives but subjective arousal ratings were not. Results suggest that stylistic expertise can be independent from general musical expertise and that stylistic expertise can contribute strongly to a listener's perception. We suggest that incorporating stylistic expertise into models of music cognition can account for differences in listeners' performance scores and can thus make models of musical cognition more robust.

Music Identification from Harmony

Ivan Jimenez, Tuire Kuusi

***Abstract*—**Earlier studies seem to indicate that harmony plays a relatively unimportant role in music identification tasks. However, identifying music from its harmony, although a difficult task, can inform our understanding of how listeners mentally represent and remember harmony. The present study is the first to examine the ability of listeners with varying musical expertise to identify well-known classical and pop/rock pieces from their chords. In this study, well-known classical and pop/rock pieces selected in a pilot were used instead of traditional tunes because their chord progressions tend to be more distinctive and their different renditions often preserve the original harmony with little or no modifications. One hundred participants, divided into four groups according to musical expertise, listened to chord progressions of representative passages from 12 pieces and were asked to identify them. Two ways of modifying pitch information in the chord progressions were explored: (1) The progressions were played with piano tones and shepard tones, the latter diminishing the effect of melodic cues. (2) The progressions were played on six transpositional levels. In the second part of the experiment, the participants heard commercial recordings of the same pieces and were asked to identify them. The study showed that music identification from harmony in an open-set task is possible, even when melodic and rhythmic cues are largely missing. Additionally, both the timbre with which the chord progressions were played and the expertise of the participants were statistically significant factors for identification; the professionals being best. We found that stimuli using shepard tones were more difficult to identify than those using piano tones, and that the major part of the difference was explained by the downgrading of melodic cues, not by timbral associations. The differences between transpositions were not statistically significant.

***Keywords*—**harmony, memory for harmony, music memory, tune identification.

I. Introduction

WHEREAS the effects of pitch and rhythmic information on tune identification have been studied extensively [1], the effect of harmony for tune identification has seldom been examined [2], [3]. There are several reasons that could explain why the effect of harmony on tune identification has not been studied more often. First, listeners can easily identify a song or piece of music based on the pitch and rhythmic features of its melody alone, without hearing its harmonic accompaniment [1]. Second, the chords that accompany a well-known melody can be changed without altering the identity of a tune [2]. Third, many tunes imply or are typically accompanied by the same or similar chord progressions, and this one-to-many correspondence can weaken the association between a chord progression and a specific tune [4], [5]. Additionally, it is difficult to measure the effect of harmony independent of the effect of melody and rhythm since a chord progression cannot be instantiated without also creating some degree of melodic motion and rhythmic activity. Reference [3] found that adding chords to a single-pitch rhythmic version of well-known tunes made identification more difficult, and it is likely that the melodic line created by the highest notes of each of the chords in the progression provided participants with a wrong melody that misled them in their process of tune identification. One way to avoid the problem of incorrect melodic cues is to voice the chords in such a way that the highest note of each chord in the progression corresponds to the most representative pitch of the melody during that time-span. The use of this type of harmonic reductions to summarize the underlying structure of music dates at least as far back as the mid-19th century [6], and formal musical training usually involves the study of harmonic reductions [7], [8]. Empirical evidence also suggests that listeners can distinguish between correct and incorrect harmonic reductions of a musical passage [9], [10]. Until the present study, however, the ability to identify music from harmonic reductions has not been empirically tested.

Ivan Jimenez is with the Department of Music, University of Pittsburgh, Pittsburgh, PA 15260 USA and the Sibelius Academy, University of the Arts, Helsinki, Finland (e-mail: iej2@pitt.edu).

Tuire Kuusi is with the Sibelius Academy, University of the Arts, Helsinki, Finland (e-mail: tuire.kuusi@uniarts.fi).

II. Pilot Study

Forty pieces, 20 popular songs and 20 pieces of classical music, were selected for the pilot based on their inclusion in recent publications: studies on tune identification [11]-[13], a study on implicit absolute pitch [14], a corpus analysis [15], a CD compilation of popular classical pieces [16], as well as because of their popularity among undergraduate music theory students at the University of Pittsburgh, who were surveyed between the years 2010 and 2014. Classical and pop/rock pieces were used instead of traditional tunes because their chord progressions tend to be more distinctive and their different renditions often preserve the original harmony with little or no modifications. In the pilot study, we tested 18 participants' ability to identify the 40 pieces from harmonic reductions played with a digital grand piano sound. Harmonic reductions consisted of six, seven, or eight chords from the initial phrase, initial period, or another representative section of the pieces and every chord lasted 2 seconds. Most harmonic reductions had four voices, and all voices had only one note per chord. Since the excerpts chosen for this pilot were the initial phrases of relatively simple pieces and songs, durational accents provided a straightforward criterion for choosing most of the notes. Harmonic reductions preserved both the contour of the different voices and chordal inversions from the original excerpts. Pitches were verified using Sonic Visualiser software. The test was administered online.

After listening to each harmonic reduction, listeners were asked to select the name of the piece from a list of the 40 pieces. The difficulty of the identification task and the high number of items in the list reduced the likelihood that listeners would identify pieces by process of elimination. The twelve pieces that were most often identified in the pilot study were selected for the main experiment: The Animals, *House of the Rising Sun;* The Beatles, *Let it Be;* Capital Cities, *Safe and Sound;* Coldplay, *Clocks;* Coldplay, *Viva La Vida;* Daft Punk, *Get Lucky;* Elgar, *Pomp and Circumstance*; Led Zeppelin, *Stairway to Heaven;* Nirvana, *Smells like Teen Spirit;* Pachelbel, *Canon in D Major;* Red Hot Chili Peppers, *Snow (Hey Oh);* Tchaikovsky*, Dance of the Sugar Plum Fairy.*

III. Main Experiment

A. Participants

There were one hundred participants in this experiment. Two participants were rejected since they had very limited familiarity with Western popular music. Eighty-seven participants were undergraduate students enrolled in one or more music classes at the University of Pittsburgh. Of the other eleven participants, five were composition doctoral or master students, two were music theorists, two were instrument instructors, one was a professional musician and one was an amateur musician. The participants were divided into four groups according to their musical background. In the first group, labeled *professionals* (N = 17; 13 male), the participants were professional musicians or professional music students. In the second group, labeled *serious amateur musicians* (N = 16; 7 male), the participants had no professional musical training but had had private instrument lessons for more than five years. The third group, labeled *amateur musicians* (N = 40; 23 male), consisted of participants with no professional musical training but who had had 5 years or less of private instrument lessons or had played an instrument for more than 5 years. The participants in the fourth group, labeled *non-musicians* (N = 25, 15 male), had not studied music nor did they actively play any instrument.

B. Stimuli

The chord sequences were first composed using Finale 2007 software. The harmonic reductions were composed so that each new harmony of the original commercial recording was composed into a chord. The specific notes for each chord in the reductions were chosen according to the same criteria used in the pilot experiment. For the nine harmonic reductions taken from popular songs, the duration of the chords matched the original duration in the most viewed version of the song on YouTube. For the three harmonic reductions taken from classical pieces, the duration of the chords was determined by averaging the duration of every chord in the ten most viewed versions of the piece on YouTube.

Both piano tones and shepard tones were used in the harmonic reductions. Shepard tones are vague in terms of pitch register, which greatly reduces the clarity of melodic gestures, voicing, and chordal inversions. For this reason shepard tones have been used regularly in experiments where researchers want to minimize the effects of melodic cues in tasks that aim to examine participants' responses to harmony [17]-[21]. Audio excerpts for the second part of the experiment were extracted from commercial recordings of the twelve pieces used in the experiment. Excerpts lasted 15 seconds and contained the harmonic reductions used in the first part of the experiment. Most excerpts from songs included vocals but no excerpt contained words from the title of the song

C. Transposition

Although most listeners can identify familiar melodies when transposed to a different key based on relative pitch information (e.g., Happy Birthday), we predicted that for pieces that are usually performed at the same pitch level, the presence of absolute pitch cues (i.e., hearing the music at its most typical pitch level) would be an important factor in the identification of a piece from harmonic reduction, that is, in the absence of important rhythmic, timbral, and textural cues. Thus, we expected that playing harmonic reductions in atypical keys would make identification more difficult. Six different pitch levels were used for the main experiment: original, one semitone down, one semitone up, two semitones down, two semitones up, and tritone.

D. Procedure

In the first part of the experiment, the 12 harmonic reductions representing the 12 pieces were presented using either piano tones or shepard tones. After hearing each reduction, the participants were asked to respond either with the name of the piece, some words of the lyrics or some description of the piece if they could not recall the exact name or lyrics. In the second part of the experiment, the tunes were presented as commercial recordings, and the participants were asked to name the pieces as well. Participants were also asked how often they had heard the piece or if they had played the piece themselves.

E. Results

The scores were summarized for each timbre and each group of participants. If a participant did not identify a piece from the commercial recording, all his/her responses for harmonic reductions of the piece were omitted from analysis. There were a few instances where participants mentioned a piece that was not in our list but that matched the chord progressions featured in the harmonic reduction. Based on that scarcity and the fact that our harmonic reductions had a less direct musical connection to those pieces than to the pieces in our list, we decided to exclude those pieces from our main analysis. Fig.1 shows the percentages of identified harmonic reductions played using two timbres for the four groups of participants. The figure shows that the professionals identified more pieces from their harmonic reductions than did the other participants.

The difference between professionals and other participants was relatively large, but the differences between the three other subgroups were small, especially when the reductions were played using shepard tones. As can also be seen, the reductions played using shepard tones were generally more difficult to identify than the reductions played using piano tones. A two-factor ANOVA analysis was made using two timbres of the reductions and the musical background of the participants as experimental variables. The analysis confirmed the results: both the timbre ($F(1, 97) = 10.031$, $p = .002$) and the musical training of participants ($F(3, 291) = 24.798$, $p < .001$) were statistically significant factors in explaining the responses. The group * timbre interaction was not statistically significant ($p = .545$). The Bonferroni post-hoc test confirmed that the subgroup of professional musicians differed from other participant subgroups ($p < .001$) and that there was no difference between the other participant subgroups. Since the subgroup of professionals differed from the other participant subgroups, and since the responses of the other three subgroups did not differ, the three subgroups were merged into one group of non-professionals, and the rest of the analyses were made using data from only two groups of participants labeled as professionals and non-professionals.

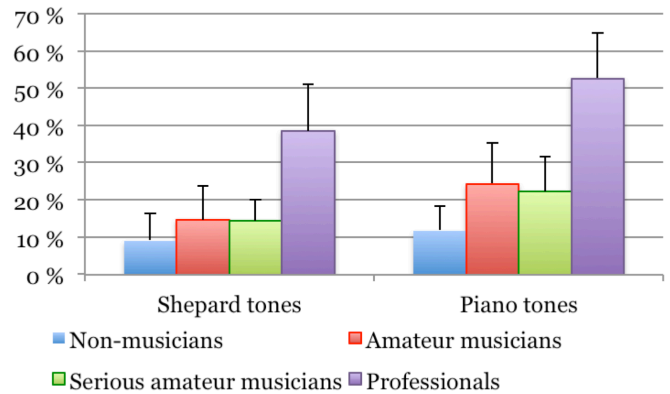


Fig. 1 Responses from four participant subgroups. Identification of music from harmonic reductions. Error bars show standard deviations

Identification rates for piano-tone and shepard-tone chordal reductions were the most dissimilar for *The House of the Rising Sun* and *Pomp and Circumstance*. The piano-tone version of the chords of *The House of the Rising Sun* was identified more often (57%) than its shepard-tone counterpart (21%). The advantage of piano over shepard tones in the case of *The House of the Rising Sun* is likely a consequence of a prominent 8-7-#6-6 melodic line that can be heard in the shepard-tone version, which is not a prominent gesture in the original song or the piano reduction. More evidence of the importance of the minimization of melodic cues in shepard-tone reductions comes from the identification of *Pomp and Circumstance* (61% for piano first and 4% for shepard first). The shepard-tone version of *Pomp and Circumstance* was most often mistakenly identified as being Pachelbel's Canon, while the piano-tone version was not. The mistake was not a consequence of the "church" association triggered by the organ-like sound of shepard tones, if that were the case, such an association would have also made Pachelbel's reduction more easily identifiable from Shepard tones than piano tones, which was not the case. A more plausible explanation for why our participants heard Pachelbel's reduction in the shepard-tone version of *Pomp and Circumstance* but not its piano-tone version is that in the shepard-tone version, the ascending bass and the chordal inversion of the original are most likely lost, and listeners are likely to hear instead the roots of the chords that create a pattern similar to that of the bass in Pachelbel's Canon.

The effect of transposition on music identification was also analyzed. Fig 2. shows the percentages of pieces identified form the six transpositional levels, done separately for professionals and non-professionals. ANOVA analysis showed that the effect of transpositions was statistically significant neither for the non-professionals ($F(5,400) = 1.127$, $p < .345$) nor for the professionals ($F(5,80) = 1.750$, $p = .131$).

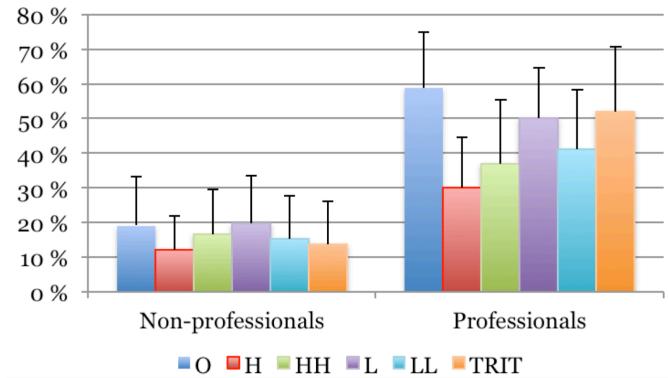


Fig 2. Effect of transposition. Correct identifications (vertical axis). O = original transposition; H = one semitone up; HH = two semitones up; L = one semitone down; LL = two semitones down; TRIT = tritone. Error bars show standard deviations.

The role of rhythmic resemblance between the original piece and the harmonic reduction on music identification was also examined. Although rhythmic resemblance is a complex topic because there are many possible considerations [22]-[24], we opted for an approach to calculating rhythmic similarity that only takes into account rhythmic density and syncopation, two aspects of rhythm that exhibited a high degree of variability in our group of 12 pieces. We calculated correlations between rhythmic similarity and the average percentages of hits vs. misses for the two participant subgroups, separately for shepard tones and piano tones, and found that there was a statistically significant correlation for both participant subgroups in the piano-tones condition ($R(12) = .759$, $p = .004$ for professionals; $R(12) = .641$, $p = .025$ for non-professionals), but not in the shepard-tone condition ($R(12) = .567$, $p = .055$ for professionals; $R(12) = .478$, $p = .116$ for non-professionals). The responses for piano-tone stimuli are consistent with the idea that rhythmic similarity does indeed facilitate identification from harmonic reductions.

IV. Discussion

Results suggest that listeners with extensive musical training are significantly better than other listeners at identifying music from its harmony. These results are in line with previous research showing that musical training increases sensitivity to harmony [25]-[29]. Our study, however, showed that neither previous musical training nor playing an instrument was necessary for identifying pieces from harmonic reductions. Results from our experiment also provide evidence of the crucial role of melodic and rhythmic cues in the identification of pieces from harmonic reductions. Pachelbel's *Canon in D* was by far the most often identified piece from harmonic reduction in our study. Although its extra-musical associations (e.g., weddings, graduations, Christmas), pervasive presence in popular culture, and distinctive sequential pattern may have facilitated its identification, the rhythmic resemblance between the block-chords in our stimuli and the beginning of Pachelbel's piece is likely to have simplified the process of identification as well. It is important to note that for most of the pieces we selected for our experiment, deviation from the slow, regular rhythm of our block-chord stimuli also meant deviation in terms of melodic content. That is, the more rhythmically dense the original excerpts were, the more melodic gestures were omitted in the harmonic reduction. The correlations between rhythmic similarity and identification rates for individual pieces from piano chordal reductions are consistent with the idea that rhythmic similarity does indeed facilitate identification from harmonic reductions. The role that melodic and rhythmic cues played in the identification of music from harmonic reductions in the present study provides further evidence of their influence on the mental representation of harmonic activity.

In addition to testing the influence of musical training in the identification of music from harmony, our experiment also tested the influence of transposition. Although music was most often identified from harmonic reductions played at their typical pitch level in our experiment, we found no significant effect of transposition on music identification.

V. Conclusions and Future Work

This study showed that music identification from harmonic reductions is possible, yet there are many factors affecting the identification process. It appears that the use of harmonic information in identification is easiest for professional musicians, especially composers and theorists. Melodic and rhythmic cues are important, and they are always present in chord progressions and harmonic reductions. The association of a harmonic reduction with a specific song is just one way that the interaction of harmony and episodic memory can influence our perception of music, suggesting that the seemingly idiosyncratic task of explicitly identifying music from harmonic reductions can improve our general understanding of how listeners make sense of music. Future research on music identification from harmony is needed to clarify whether the association of harmony with episodic memory occurs implicitly or not as well as how such unconscious associations impact our experiences with music.

Acknowledgment

The Authors would like to thank Matti Strahlendorff for the shepard tone stimuli used in this study and Lynne Sunderman for her proofreading of this manuscript.

References

[1] A. R. Halpern and J. C. Bartlett, "Memory for melodies." in *Music perception*. Springer, New York, 2010, pp. 233-258.

[2] D.J. Povel and R. Van Egmond, "The function of accompanying chords in the recognition of melodic fragments," *Music Perception,* vol. 11, pp. 101-115, 1993.

[3] T. Kuusi, "Tune recognition from melody, rhythm and harmony," in *Proc. 7th Triennial Conference of European Society for the Cognitive Sciences of Music (ESCOM),* Jyväskylä, Finland. 2009.

[4] R. J. Scott, *Money Chords: A Songwriter's Sourcebook of Popular Chord Progressions*. Bloomington, IN: iUniverse, 2000.

[5] N. Stoia, "The common stock of schemes in early blues and country music," *Music Theory Spectrum,* vol. 35, pp. 194-234, 2013.

[6] C. Czerny, *School of Practical Composition: Complete Treatise on the Composition of All Kinds of Music* (J. Beshop, Trans.). New York: Da Capo Press, 1979 [Reprinted from *Die Schule der Praktischen Tonsetzkunst*, Bonn: Simrock, 1849].

[7] E. Aldwell, C. Schachter, and A. C. Cadwallader, *Harmony & Voice Leading*, 4th ed. Boston, MA: Schirmer/Cengage Learning, 2011.

[8] J. P. Clendinning, and E. W. Marvin, *The Musician's Guide to Theory and Analysis*, 2nd ed. New York: W.W. Norton, 2011.

[9] M. L. Serafine, N. Glassman, and C. Overbeeke, "The cognitive reality of hierarchic structure in music," *Music Perception,* vol. 6, pp. 397-430, 1989.

[10] N. Dibben, "The cognitive reality of hierarchic structure in tonal and atonal music," *Music Perception,* vol. 12, pp. 1-25, 1994.

[11] C. L. Krumhansl, "Plink: 'thin slices' of music," *Music Perception*, vol. 27, no. 5, pp. 337-354, 2010.

[12] K. VanWeelden, "Classical music as popular music: adolescents' recognition of Western art music." *Applications of Research in Music Education,* vol. 31, pp. 1–11, 2012.

[13] K. VanWeelden, "Preservice music teachers' recognition of Western art music found in popular culture. *Journal of Music Teacher Education,* vol. 24, pp. 65–75, 2014.

[14] K. Frieler, T. Fischinger, K. Schlemmer, K. Lothwesen, K. Jakubowski and D. Müllensiefen, "Absolute memory for pitch: A comparative replication of Levitin's 1994 study in six European labs," *Musicae Scientiae,* vol. 17, pp. 334-349, 2013.

[15] T. deClercq and D. Temperley, "A corpus analysis of rock harmony," *Popular Music*, vol. 30, pp. 47–70, 2011.

[16] D. Parry, *The 50 Greatest Pieces of Classical Music*. London Philharmonic Orchestra. X5 Music Group. ASIN: B002WMP162, 2009.

[17] J. J. Bharucha, "Anchoring effects in music: The resolution of dissonance," *Cognitive Psychology,* vol. 16, pp. 485-518, 1984.

[18] J. J. Bharucha and K. Stoeckig, "Reaction time and musical expectancy: Priming of chords." *Journal of Experimental Psychology: Human Perception and Performance*, vol. 12, pp. 403-410, 1986.

[19] É. A. Firmino, J. L. O. Bueno, and E. Bigand, "Travelling through pitch space speeds up musical time," *Music Perception, vol. 26*, pp. 205–209, 2009.

[20] T. Kuusi, "Interval-class and order of presentation affect interval discrimination," *Journal of New Music Research,* vol. 36, pp. 95–104, 2007.

[21] C. L. Krumhansl, J. J. Bharucha, and E. J. Kessler. "Perceived harmonic structure of chords in three related musical keys," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 8, pp. 24-36, 1982.

[22] E. Cao, M. Lotstein, and P. N. Johnson-Laird, "Similarity and Families of Musical Rhythms," *Music Perception,* vol. 31, pp. 444-469, 2014.

[23] J. Forth, *Cognitively-motivated Geometric Methods of Pattern Discovery and Models of Similarity in Music*. PhD Dissertation, Goldsmiths University London, 2012.

[24] G. T. Toussaint, *The Geometry of Musical Rhythm: What Makes a "Good" Rhythm Good?*. Boca Raton, FL: CRC Press, Taylor & Francis Group, 2013.

[25] M. M. Farbood, "A parametric, temporal model of musical tension," *Music Perception*, vol. 29, pp. 387–428, 2012.

[26] R. Kopiez and F. Platz. "The role of listening expertise, attention, and musical style in the perception of clash of keys," *Music Perception,* vol. 26, no. 4, pp. 321-334, 2009.

[27] D. Sears, W. E. Caplin, and S. McAdams,"Perceiving the Classical cadence." *Music Perception, vol*. 31, no. 5, pp. 397-417, 2014.

[28] L. R. Williams, *The Effect of Musical Training and Musical Complexity on Focus of Attention to Melody or Harmony*. Unpublished doctoral dissertation, Florida State University, Tallahassee, FL., 2004.

[29] R. S. Wolpert, "Attention to key in a non-directed music listening task: Musicians vs. nonmusicians," *Music Perception*, vol. 18, pp. 225–230, 2000.

Memory for the Meaningless: Expert Musicians' Big Advantage at Recalling Unstructured Material

Giovanni Sala[1], Fernand Gobet[1]
[1]Department of Psychology, University of Liverpool, Liverpool, UK

When recalling sequences of notes briefly presented in a musical stave, expert musicians show a better performance then novices. Interestingly, this skill effect occurs even with unstructured material – such as random sequences of notes. Classically, the advantage of experts at recalling domain-specific meaningless material – such as shuffled chess positions, or random notes – has been convincingly explained by the chunking theory. In fact, theories assuming small memory structures (i.e., chunks) predict a *moderate* skill effect, because even in unstructured material some small meaningful structure – which experts are more likely to recognize and recall than novices – can occur by chance.

The aim of this meta-analysis is thus to establish (a) the *size* of experts' superiority at recalling unstructured domain-specific material, and (b) the possible differences between diverse domains of expertise (music, programming, board games, and sports). A systematic search strategy is used to find studies comparing experts' and novices' performance at recalling random material. Then, a random meta-analytic model (K = 28) is run to calculate the overall correlation between expertise and performance. The results show (a) a moderate overall correlation ($\bar{r}$ = .332, 95% CI [.221; .434], p < .001), and surprisingly (b) a significantly superior correlation ($\bar{r}$ = .706, 95% CI [.539; .820], b = 0.455, z = 2.54, p = .011) in the domain of music compared to all the other domains considered.

These outcomes suggest that expert memory is qualitatively different in music, because such a large skill effect cannot be accounted for only by meaningful chunks of notes occurring by chance. Possibly, whilst experts in other fields base their superiority on a greater amount of small meaningful domain-specific structures stored in their long-term memory, expert musicians benefit from a retrieval structure specific for notes drawn in the musical stave, and able to explain such a strong correlation.

Memory for melody: How previous knowledge shapes the formation of new memories

Steffen A. Herff[1], Kirk N. Olsen[1] and Roger T. Dean[1]
[1]MARCS Institute for Brain, Behaviour & Development, Western Sydney University, Sydney, NSW, Australia

Most memory domains show clear performance decrements as the number of intervening items between a stimulus' first exposure and its reoccurrence increases. For example, spoken word recognition decreases as the number of additional intervening words increases between a target word's first and second occurrence. However, using continuous recognition paradigms in the context of Western tonal music, our work demonstrates that this is not the case in memory for melody: the first intervening melody results in a significant performance decrement; however, every additional intervening melody does not impede melody recognition. We have replicated this observation multiple times using different melody corpora, implicit/explicit measurements, and up to 190 intervening melodies. To explain these results, we posit a model describing memory for melody as the formation of multiple representations that can regenerate one another; specifically, a representation of the melody as an integrated whole, and a representation of its underlying parts. The model predicts that removing a familiar musical context established by previous musical experience disrupts integration of melodies as a whole. In this case, decrements in recognition performance should be observed as the number of intervening melodies increases. This hypothesis was tested in the present study using novel melodies in a microtonal tuning system unfamiliar to participants. The study used the same continuous recognition paradigm that previously produced no effect of the number of intervening items in response to Western tonal stimuli. Results provide preliminary support for the model's predictions and show that when presented with an unfamiliar tonal system, memory for melodies does indeed deteriorate as the number of intervening items continuously increases beyond immediate repetition. Overall, the results advance our understanding of memory mechanisms and pave the way for future work designed to predict melody recognition.

Generality of the memory advantage for vocal melodies

Michael W. Weiss[1], E. Glenn Schellenberg[1], and Sandra E. Trehub[1]
[1]Department of Psychology, University of Toronto Mississauga, Mississauga, Ontario, Canada

Adults (musicians and nonmusicians) and older children exhibit better memory for vocal melodies (without lyrics) than for instrumental melodies (Weiss et al., 2012, 2015a, b), which implies processing advantages for a biologically significant timbre. Because previous studies used a single female vocalist, it is possible that something about her voice or about female voices in general facilitated melodic processing and memory. In the present study, we compared adults' memory for vocal and instrumental melodies, as before, but with two additional singers, one female (same pitch level as the original female) and one male (7 semitones lower). Instrumental melodies were MIDI versions at the same pitch level as the corresponding vocal recording (i.e., male or female). (Previous research showed no memory differences for natural and MIDI versions.) During the exposure phase, participants rated their liking of 24 melodies (6 each in voice, piano, banjo, and marimba), with random assignment to one of three voices (n = 29, female from previous research; n = 32, new female; n = 29, male). After a short break, recognition memory was tested for the same melodies plus 24 timbre-matched foils (6 per timbre). A mixed-model ANOVA and follow-up pairwise comparisons revealed better recognition memory for vocal melodies than for melodies in every other timbre (ps < .001), replicating previous findings. Importantly, the memory advantage was comparable across voices (i.e., no interaction, $F < 1$) despite the fact that liking ratings for vocal melodies differed by singer. Our results provide support for the notion that the vocal advantage in memory for melodies is independent of the idiosyncrasies of specific singers and of vocal attractiveness, arising instead from heightened arousal or attention to a biologically significant timbre.

Musical Experience and Executive Function

Brooke M. Okada[1], L. Robert Slevc[1]
[1]Psychology, University of Maryland, College Park, MD, USA

Learning and performing music draws on a host of cognitive abilities; might musicians then have advantages in related cognitive processes? One likely aspect of cognition that may be related to musical training is executive function (EF). EF is a set of top-down processes that regulate behavior and cognition according to task demands (Miyake et al., 2000). Models of EF postulate three related, but separable components: inhibition, used to override a prepotent response; shifting, used to switch between task demands; and updating, used to constantly add and delete items from working memory.

Musicians likely draw on EF; for example, using inhibition to resolve musical ambiguities, shifting to coordinate attention when playing with others, and updating when sight-reading music. Because of this, many studies have investigated whether music training is related to more general EF advantages. However, results from such studies are mixed and difficult to compare. In part, this is because most studies look at only one specific cognitive process, and even studies looking at the same process use different experimental tasks.

The current study addresses these issues by investigating how individual differences in all three EF components relate to musical training. By administering a well-validated EF battery of multiple tasks tapping each EF component and a comprehensive measure of musical training and sophistication (Müllensiefen et al., 2014), we can obtain reliable measures of individual differences in EF and musical experience.

Preliminary analyses (N=72, planned N=150) suggest that, although musical sophistication correlates with some EF tasks, these relationships do not persist after controlling for the potential confounds of general intelligence, socio-economic status, and handedness. These results suggest that observed relationships between musical training and EF may instead reflect underlying differences in other factors such as IQ or SES.

Divided attention and expertise in the context of continuous melody recognition

Steffen A. Herff[1,2] and Daniela Czernochowski[2,3]
[1]MARCS Institute for Brain, Behaviour & Development, Western Sydney University, Sydney, NSW, Australia
[2]Institut für Experimentelle Psychologie, Heinrich Heine Universität Düsseldorf, Düsseldorf, NRW, Germany
[3]Cognitive Science, Technische Universität Kaiserslautern, Kaiserslautern, Rheinland-Pfalz, Germany

Allocating attention to different tasks is an everyday requirement that has a significant impact on human memory. The popular view of resource models suggests that a finite attentional resource is assigned to different tasks. If only one task is present, attention converges around that task, but as soon as a secondary task is to be completed simultaneously, attention needs to be divided. Memory performance tends to suffer as fewer resources are allocated to a task. However, this performance decrement is shaped by domain specific expertise. The present study investigates how musical expertise shapes the influence of divided attention on memory for melodies. Musicians, Non-Musicians and intermediate Novices were recruited to perform a melody recognition task in a continuous recognition paradigm. Participants were randomly assigned to an undivided or to a divided attention condition, solving a digit-monitoring task simultaneously. Musicians outperformed Non-Musicians in the melody recognition task, with Novices falling between the other two groups. A decrease in absolute recognition performance was statistically identical in all expertise groups in the divided attention condition compared to the undivided attention condition. However, despite no inherent advantage of Musicians in digit monitoring tasks, musicians performed significantly better in the digit-monitoring task than Non-Musicians during the divided attention condition. Results are discussed in the light of attentional resource models and provide insight on the interaction of expertise, attention and memory. An augmented resource model is suggested to explain the results, theorizing an asymmetrical, non-linear trade off between two simultaneous tasks. The model assumes that an identical absolute performance decrement at the high end of the performance spectrum releases more attentional resources that can be allocated to another task, compared to the lower end of the performance spectrum.

Generalizing the learning of instrument identities across pitch registers

Stephen McAdams[1], Ayla Tse[2], Grace Wang[3]
[1]Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT), Schulich School of Music, McGill University, Montreal, QC, CANADA
[2]Department of Psychology, McGill University, Montreal, QC, CANADA
[3]Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT), Cognitive Science Program, McGill University, Montreal, QC, CANADA

The notion of "a musical instrument's timbre" is an unfortunate misconception that confuses source identification with the perception of the attribute timbre, which covaries with pitch and dynamics in mechanical instruments. Research has shown that sounds separated by more than an octave are often not recognized by nonmusicians as being produced by the same sound source. To study how listeners construct the representation of a single sound source from a set of sounds that vary in timbre across pitch, we tested how nonmusicians learned to identify instruments. The hypothesis was that the presentation of information concerning how pitch and timbre covary would increase the ability to generalize identification across pitch registers. In Exp. 1, participants first learned to associate the names of 11 instruments from the woodwind, brass, pitched percussion, and plucked, bowed, and struck string families with the sounds produced by these instruments at a pitch of C4 and dynamic of *mf*. In Exp. 2, the training stimuli consisted of a diminished arpeggio from G3 to G4, thus exposing the listeners to the way timbre varied with pitch for each instrument within an octave. Training was completed once performance reached a minimum of 75% correct over 4 successive blocks of the 11 sounds within a maximum of 20 blocks. Half of the participants failed to reach criterion performance in Experiment 1, but all did so in Experiment 2. Successful participants were then tested on their ability to generalize identification from the learned sound(s) to single notes across each instrument's full range. Performance varied significantly across instruments: some generalized to a greater extent than others. The results are very similar for the two experiments, suggesting that there is no gain in generalizing identification from the multi-pitch condition to the single-pitch condition. These results will be discussed in terms of the acquisition of models of musical sound source identity.

Playing by Ear beyond 100: A Continuing Case Study

Eleanor Selfridge-Field

***Abstract*—**The subject ME, who was introduced at the age of 95 in a study presented at ICMPC 11, retains performance abilities at 101 that astound many listeners. Vascular dementia is held to account for a loss of memory for biographical detail and some spatial-recognition skills. Her performance of musical tasks and her ability to learn new music on aural exposure have survived a recent stroke and other physical maladies intact. This update provides additional data on health impacts, trends in her recent performances, additional information on her background, and consideration of recent studies.

***Keywords*—**vascular (non-Alzheimer's) dementia, lexical memory, semantic memory, executive function, playing by ear (pertinent skills).

I. Introduction

THE continuing survival of the subject ME, introduced in "Playing by Ear at 95" (ICMPC 11), has enabled us to collect six years of data pertinent to the preservation of her performance abilities. Playing by ear requires a strong semantic sense of musical structure in conjunction with other skills. Retention of all of them is essential to preserving this capability.

ME was born into a highly literate family in eastern Tennessee in August 1914 [1]. Further investigation of her background has added flesh to the skeleton previously provided. At the end of the First World War (1918) her family moved to the Atlanta area, where one set of grandparents lived. In their home ME encountered the piano-playing of a 6-year-old cousin and soon tried to emulate it. She attributes to this cousin the explanation she still gives today of her approach to playing by ear.

She started piano and violin lessons while the family was still in Georgia. Her family returned to Tennessee, settling in the Chattanooga area, by the time she reached junior high school (c.1926). There she enrolled in a music school started by a Czech immigrant in 1909. It offered classes, ensembles, and individual lessons. ME studied a number of instruments, was involved in ensembles of various kinds, and sometimes instructed younger children. She earned degrees in music education in Murfreesboro (at a predecessor of Middle Tennessee University) and then in Nashville. Her aim had been to become a band and orchestra teacher, but emerging in the middle of the depression, she failed to find a teaching position. After a brief stint as a rural social worker, she took a position as violist in a women's orchestra in Nashville. She occasionally played trombone as well.

ME was married in 1940. She and her husband welcomed two sons over the next few years. In 1946 the family moved to Jacksonville (FL). There she engaged in no substantial musical activities utilizing her earlier skills until she was in her 90s, but informal piano playing was a routine activity at holiday gatherings. After their children were grown, ME and her spouse took up square dancing and visited other square-dance venues on regular trips across the US to visit their sons. She also played in a hand-bell choir in her retirement years.

In her 80s ME had a TIA, but its effects were so slight that no extensive medical examination was conducted. Only when she developed signs of dementia in her early nineties did her sons move her to an assisted-living residence in Florida. There a kindly director, discovering that she played the piano, encouraged her to give short programs for other residents. This has been her main source of musical engagement during the intervening years, particularly since her move to California (2008). On December 4, 2014 (less than four months after her 100th birthday) ME suffered a stroke. Below we review her musical activities over the past six years.

A. Performing venues and circumstances

In California ME has played in a total of seven venues. She visits two of them (both assisted living facilities much like her own) on a regular monthly basis. She plays for holidays in her own residence and in her son's church. She performs from time to time in the skilled nursing home to which she is sent when requiring extended care. In warm weather her son sometimes takes her to a nearby park with an accessible piano.

No maintenance is provided for any of the pianos. Her audiences consist mainly of residents and staff members in the homes. Such employees are appreciative of her skills and of the fact that at 101 she is still an active volunteer. Two weeks after her recent stroke ME was asked by a nurse in her extended care facility if she would play some carols for the staff holiday party. She happily complied.

In one home a resident clarinetist (MC, formerly a professional musician) sometimes plays with her. Other residents may sing or dance on a whim, but MC, who is blind and has breathing problems, is the only one who frequently joins her in performances.

Eleanor Selfridge-Field is a consulting professor of music at Stanford University, Braun Music Center #130, 541 Lasuen Mall, Stanford, CA 94305-3076, USA (esfield@stanford.edu). She conducts digital research at the Center for Computer Assisted Research in the Humanities, an affiliate of the Packard Humanities Institute (Los Altos, CA 94022).

II. Musical Resources for Analysis

Data have been gathered in several ways. We have a video (1990) of a family gathering in Tennessee in which several family members of ME's generation played the piano. One also sang. We have a one-hour Disklavier recording from Dec. 2009, with extracted MIDI files that were discussed in [1]. Over the past six years ME's family has made 16 audio recordings of her appearances in the two homes and the church. These typically contain 30 minutes of music. When possible, direct observation has been made.

A. Repertory overview

One of ME's sons began compiling a list of pieces she "remembered" when she first entered assisted living. The director wanted some way of asking for what was possible. Additions continue to be made, as requests that she has satisfied are added. (She can almost always satisfy an ad hoc request.) The database currently contains almost 400 pieces, and it is clear that ME can in principle play many hundreds more, since her performances always amount to a fresh reconstruction of something aurally familiar. One value of the growing database is that at each appearance she is rotated to a new starting position on the virtual dial. A list of titles (as many as 30) is usually prepared in advance.

B. Peak acquisition period

While it is not possible to determine when any given piece of music first entered the recall lexicon of musicians who play by ear, we have used dates of first publication to give a sketchy profile of the periods during which the piece was well known. At present 315 pieces have been anchored to such a year. These come with the caveat that sheet-music publication rises and falls with economic contingencies. Many pieces in ME's repertory were reissued in new arrangements from era to era.

TABLE I
Chronological Distribution of Repertory

Time span	No.	Percentage
1586-1860	23	7.4%
1861-1880	13	4.1%
1880-1900	15	4.7%
1901-1910	20	6.3%
1911-1920	36	11.5%
1921-1930	81	25.7%
1931-1940	58	18.4%
1941-1950	26	8.2%
1951-1960	25	7.9%
1961-1980	6	1.9%
1981-2000	2	0.7%
2001-2016	10	3.2%

Dates indicate year of original publication.

Fifty-five titles first appeared before ME was born. The overall year range is 1586-2016. (ME continues to learn new music.) In the distribution of publication dates (TABLE I), one can see a rapid decline in the acquisition of new titles between 1961 and 2000. Note that her rate of acquiring new material in recent years has modestly increased, presumably in response to greater musical engagement. This distribution also shows that the bulk of ME's performing repertory comes from the decades in which she was most active musically (to 1940), but because of the likelihood that she learned some titles later than indicated here, her learning curve may peak slightly later. The musical culture in which ME was immersed was a live one in which nearly everyone with whom she associated made music.

C. Genres and tonal properties

ME's repertory consists largely of popular music, but under that rubric we find instances of every kind of music in common use in southern Appalachia during ME's formative and early adult years. Of the 365 titles so far identified, popular titles number 242 (66.3%). Other genres include songs popularized in Broadway shows and films of musicals (8.2%) and hymns (6.8%). Music of a more indigenous character (ballads, folk and traditional music, barbershop, country, Dixie, gospel songs, spirituals, jazz, and swing collectively make up 13%. The remainder await classification.

With respect to tonal orientation, most pieces (82%) are in major keys. A further 7% are in minor or modal frameworks, 7% are unidentified, 3% are in "blues" scales, and a few (1%, probably Appalachian) are pentatonic.

III. Mental, Physical, and Cognitive Status

Health professionals classify ME's dementia as one of vascular origin. No suggestion of corollary Alzheimer's has been made. One of my colleagues has commented that her profile suggests amnesia. (No instance of physical trauma has been identified to substantiate this hypothesis.) Her mini-mental status exam (MMSE) score is static at 23. Her greatest strengths are in digit span and calculation generally, although the profiles of outstanding subjects in experimental work on digit-span such as [2] diverge substantially from hers in being much younger and in being tested in abstract rather than practical tasks. ME's weaknesses involve lapses in visual memory (for people, places, and regular routes of local travel). Her eyesight is sufficiently good not to require correction. Her hearing in one ear has declined, but age-adjusted she still hears very well. Her aural memory is superior.

ME's short-term memory deficiencies progress towards recency. The retrograde aspect of her memory has been static over the past six years. What she remembers is selective. Her autobiographical memory dwells in the years from ages 3 to 8. Her visual memory falters in early adulthood. She recognizes family members and those whom she has met prior to the age of 75 or 80. For her 100th birthday a sister 13 years her junior and a middle-aged niece traveled from the east coast to join the celebration. ME recognized both instantly and found the reunion deeply affecting. (ME's sister worked as a radio announcer for about 35 years.) When the local mayor came to her facility to read a birthday proclamation she appreciated the associated festivity but seemed unaware that it related to her.

ME does not hesitate to initiate conversations. She shares with many dementia patients a tendency to propose generic topics of conversation. "Beautiful day today, isn't it?" is one of her standard openers. Her most prominent social quality is a perennially cheerful disposition (possibly enhanced by short-term memory loss, since she never retains any recollection of her numerous physical disabilities).

Over the past six years ME had been hospitalized twice for pneumonia, three times for falls, and once for a stroke. A fall that resulted in several broken ribs, giving rise to a 4-month long convalescence, has probably contributed greater disability than her 2014 stroke. ME has been ordered to use a wheelchair, though she never misses an opportunity to stand in place. The stroke did not cause noticeable cognitive decline, but a concomitant advance of distal arthritis, particularly affecting the knuckles of her left hand, has sometimes interfered with her ability to play as fast as she would like. The combination of the wheelchair (which distances her from keyboard) and decreasing manual dexterity has not discouraged her from playing. Her determination and optimism compensate.

Common methods of subject investigation (including fMRI) are deemed inappropriate for a person of ME's age. Her medical doctor stands firm in his opposition to tests (mental or physical) that "do not benefit a patient's condition."

ME occasionally acknowledges her loss of short-term memory with a wry comment. Asked recently whether she heard the noise from one of Northern California's rare thunderstorms the night before, she answered "You know *I* wouldn't know the answer to a question like that."

She clearly has an emotional response to music but it is not definably autobiographical. A lack of living witnesses to her early life is an acknowledged drawback to answering some questions about her past. Her colleague MC occasionally chides her that a song she picks is "too old" for him. (MC is 12-15 years ME's junior.)

Music appears to enhance ME's cognizance of her present circumstances. She likes to accompany one singer she met on a visit to California 30 years ago. Yet she is never sure whether the place in which she is currently playing is her home or not. To find her room she relies on the nameplate on the door. She does not know whether she shares the room (she does). At the same time her uptake of new musical material on the strength of one or two hearings is still remarkable. Her daughter-in-law is responsible for exposing her to new material.

IV. Performing Qualities

In recent recordings made by ME's family and friends in recent years audio quality varies from venue to venue. All recordings contain background noise (conversation, requests, staff instruction, and in one venue the cheerful contributions of resident canaries). MC's clarinet and sundry singers are captured in combination with ME's piano playing in several instances. MC's participation cannot be anticipated. He sometimes attends ME's performance without participating.

We have used the Disklavier recordings from December 2009 [1] and their corresponding MIDI files to compare with subsequent performances of the same works, although with 400 pieces to choose from the same titles do not reappear rapidly. Because of her physical limitations, it has not been feasible to bring ME to the Disklavier in the intervening years.

Only one piece has been fully transcribed from a performance. It shows her own composition, a song called "Give me a rainbow" (September 2005). Its lyrics were written by a then nonagenarian friend of ME's. Her son's transcription defines the status of her harmonic vocabulary then and reveals elements of her rhythmic and harmonic practice in ways that audio resources cannot make a clear.

A. Harmonic scope

One of the outstanding features of ME's playing is the rich harmonic vocabulary she utilizes. It established that her chord usage is more dependent on jazz and blues idioms than on harmonic practices based on European art music. The melody out of context would suggest many fewer chords than she actually uses (Table II). Among these, various types of tonic chords are particularly numerous. Heavy emphasis is placed on various mutations of degrees I and II. The treatment of chords based on degrees IV, V, and VI is similar. Seventh chords with various alterations are plentiful. A few sixth chords and altered roots also occur. (Alternative spellings are possible in some cases.) No median root (III) is ever employed, and only a few instances of the VII root occur.

Table II
Chord Types in ME's 2005 Composition

Root C	Root D	Root F	Root G	Root A	Root B
I maj	II maj	IV maj	V maj	VI maj	
I min	II min			VI min	
I dim	II dim				
I 7		IV 7	V 7	VI 7	
I b7		IV b7	V b7	VI b7	
I #3/b7	II #3/f7	IV #3/b7	V #3/n7		
	II #3/#7				
					VII dim/ b7
I #5					
I #5/b7					
	II #5/n7				
				VI dim7	

Roman numerals indicate conventional chord degrees.

B. Metrical and rhythmic features

ME's performances are full of crisp rhythms. In successive iterations of a single piece she uses unexpected syncopations, unexpected rests, and staccato passages for variety. In multiple performances of the same pieces her tempos are not identical.

C. ME with and without other musicians

In accompanying other performers, such as the clarinetist MC, ME's ability to emphasize the harmonic framework and omit the melody (entirely by ear) seems to be instinctive. She transitions in and out of accompaniment mode without the slightest hesitation. She does not remember (unless reminded) that clarinetists prefer to play in Bb. The usual result of this lapse is that MC tends to drop out of modulatory B sections. Because of breathing difficulties, he rarely plays for more than one verse. ME in consequence is also inclined to stop after one verse when playing with MC.

In her best form ME plays three verses of a piece. No two are played the same way. Of the second and third one is usually in a high register, the other in a low one. There is no set order. She employs a wide range of articulatory methods to different each verse. Her arrangements vary somewhat by musical genre. ME has a clear sense of what is appropriate where.

When she has an engaged audience (something she can assess accurately) she plays more pieces, and plays them with more variations than she would otherwise do. Her programs are intended to last 30 minutes, but the number of works and their duration are reciprocal. If playing with fewer repetitions, she plays more works. The maximum number of pieces in her programs is 30. The average length of a single work is 1'40". Six years ago it was well above 2 minutes, but fewer titles appeared on her programs.

D. Non-imitative contrapuntal interpolations

A familiar feature of ME's playing is her filling in of intermediate cadences with melodic interjections. It is difficult to characterize them verbally, but collectively they suggest an instinctive avoidance of static harmonies and drab melodies. Considering the degree to which she emphasizes accents (evident in the spikes on sonograms) and her frequent use of sudden rests to dramatize the approach of a cadence, we can suggest that she is highly skilled at eliciting the attention of listeners. Her elderly listeners are attentive. Feet tap on wheelchair footrests even while heads droop. Fingers tap on armrests, or in the severely impaired eyelids may flicker in time to the music. Somnolent residents occasionally burst forth in song as they suddenly recognize an old favorite. There may be valuable lessons here beyond a reckoning of ME's skills.

Cognitive Aspects of Skill Retention

At 101 ME does not play with the same vigor as she did at 75. Being the consummate musician that she is, ME has developed her own methods to compensate for her disabilities. Yet over the six-year period in which we have observed her, most aspects of her playing are little changed.

She consistently retains three features: (1) a strong, steady beat, (2) an instinctive differentiation of musical foreground and background, and (3) an excellent sense of musical structure. The third may be the key to her confidence in playing for others. Additionally, she is not distracted by external auditory stimuli, including her own occasional errors.

Her single composition is in 3/4. An anticipatory introduction (which was a common feature in sheet music up to the 1960s) leads to a substantial piece in AABA form. A coda (echoing the final line) is appended in a higher register. Phrase structure is dictated by the underlying verse, which is slightly irregular. The A section consists of four 6-bar couplets, the B section (really B-B_1) of two 8-bar lines. A four-octave chromatic glissando (G2-G6) serves as a cadenza leading to the recapitulation. The total length is 110 bars.

In comparison to performances of the same piece over the past decade, this schematic detail silhouettes ME's gradual drift towards simplification of detail. Basic architecture, meter, and accent are never altered. In recent performances of this piece the introduction, cadenza, and coda have usually been omitted. When ME is tired she may abbreviate the form to AABA (excluding the introduction and coda) or even AA. When her fingers do not obey her ear, she is learning to make a virtue of necessity. We may hear tone clusters where chromatic runs used to be, but the clusters always fall on the beat and most listeners do not notice the difference.

V. Discussion

Playing by ear is widely considered to depend on musical memory, but the individual skills it requires are substantially different from those required to memorize music learned from a notated source. Recognizing this distinction is essential to appreciating the elements of ME's performances that are really special.

Playing by ear relies on a sense of structure with which ME is well endowed. It is used in combination with extensive (previously stored) lexicons of melodic and harmonic knowledge, which themselves interact with other processes that range from the general to the specific. The interplay of cognitive skills that operate along this continuum in live performance relies heavily on the synchronization of musical learning, cognitive schemata, and motor memory. Each aspect must accommodate aural stimuli as they occur. Playing by ear relies more nearly on improvisation skills than on memorization. The governing of these processes could well lie within the province of executive function.

One model that may be concordant with the cognitive model we describe in ME comes from recent work by Omar [3], [4]. The two studies represent a total of three subjects. One has no prior musical training. Another is almost 50 years younger than ME. Unlike ME, these three subjects have semantic dementia. They retain knowledge of musical "objects" and "knowledge" but lack several other skills well preserved in ME. We could say that their losses are the inverse of ME's strength.

ME plays without premeditation. (When asked, ME likes to say that what she does is easy. She finds the starting note and her fingers do all the rest.) Her insistence that she has never read music is contradicted by her ability to sight-read choral music, photographic evidence of her participation in an orchestra, and two degrees in music. In contrast to the scholarly literature on music and dementia, anecdotal evidence of other accomplished musicians who seem to function somewhat like ME at ages above 90 are accumulating in newspapers and online videos.

One area of current research that may be worth evaluating in relation to studies of the preservation of musical function can be found in the cognitively enhanced music-theoretic literature such as the works of Temperley [5]-[6]. Another fruitful area for reverse "engineering" may exist in the explorations of audio segmentation in a series of papers by Pearce, Müllensiefen, and Wiggins [7]-[8]. They compare statistical and rule-based approaches to segmentation, such that in their melodic grouping formulae the critical features are rests, the length of attach points, changes of register, and changes of phrase length. Features 1, 3, and 4 figure prominently in ME's methods of structural articulation.

VI. Therapeutic Implications

Therapeutic investigations bring further perspectives that may be relevant. Reference [9] found that giving older subjects four months of piano instruction led to improvements in executive function, visual scanning, and motor ability. This followed studies [10]-[11] that concentrated on the value on piano instruction on executive function in adults aged 60-85.

Looking at a slightly different aspect of music in the presence of dementia Ren and Luo [12] found positive benefits in cognitive processing speed and working memory but not in fluid intelligence in elderly subjects. Playing by ear requires fluid application of learned principles, while reading from a score does not. However, reading fluently from a score relies on fluid recall of kinesthetic memory. In the Community of Voices project [13] amateur singers are recruited from senior centers in order to evaluate the benefits of musical participation without prior training.

A paucity of studies of multi-dimensional skill coordination in aged populations leaves much room for speculation and future research. Subjects such as ME provide unique case histories. At all stages of her life ME has to have been an exceptionally observant listener and a quick-witted performer. Given the near impossibility of finding matches for subjects beyond the age of 90, individual case studies seem to offer the only practical way into the *terra incognita* of musical skill retention at advanced ages.

ACKNOWLEDGMENTS

ME's family (unnamed here to protect ME's anonymity) has provided generous access and audio support. Craig Sapp has contributed valuable technical assistance. Suggestions by D. Gordon Bower, K. Anders Ericsson, Caroline Palmer, Petr Janata, Julene Johnson, Dang Vu-Phan, and Anthony Wagner have helped to steer my thinking in productive directions.

REFERENCES

[1] E. Selfridge-Field, "Playing by ear at 95: A case study," *ICMPC 11 Abstracts* (Seattle, 2011), ed. S. M. Demorest, S. J. Morrison, and P. S. Campbell, p. 91.

[2] Y. Hu and K. A. Ericsson, "Memorization and recall of very long lists accounts for within the Long-Term Working Memory framework," *Cognitive Psychology,* vol. 64 (2012), pp. 235-266.

[3] J. C. Hailstone, R. Omar, J. D. Warren, "Relatively preserved knowledge in semantic dementia," *J. Neurol. Neurosurg Psychiatry,* vol. 80/7 (2009), pp. 808-809.

[4] R. Omar, J. C. Hailstone, J. E. Warren, S. J. Crutch, and J. D. Warren, "The cognitive organization of music knowledge: a clinical analysis," *Brain,* vol. 10 (2010), pp. 1200-1213. doi: http://dx.doi.org/10.1093/brain/awp345.

[5] D. Temerpley, *The cognition of basic musical structures*, Cambridge: MIT Press, 2001.

[6] D. Temperley, *Music and probability*, Cambridge: MIT Press, 2007.

[7] M. T. Pearce, D. Müllensiefen, and G. A. Wiggins, "A comparison of statistical and rule-based models of melodic segmentation," ISMIR 2008 Proceedings, ed. J. P. Bello, E. Chew, and D. Turnbull, http://www.mirlab.org/conference_papers/International_Conference/ISMIR%202008/papers/ISMIR2008_228.pdf.

[8] M. T. Pearce, D. Müllensiefen, and G. A. Wiggins, "Melodic grouping in music information retrieval: new methods and applications," *Adv. in music inform, retrieval*, ed. Z. W. Raś and A. A. Wieczorkowska (Berlin, 2010), pp. 364-389.

[9] S. Seinfeld, H. Figueroa, J. Ortiz-Gil, and M. V. Sanchez-Vives, "Effects of music learning and piano practice practice on cognitive function, mood, and quality of life in older adults," *Frontiers in Psychology*, vol. 4 (Nov. 1, 2013). doi:http://dx.doi.org/10.3389/fpsyg.2013.00810.

[10] J. A. Bugos, "The effects of individualized piano instruction on executive functions in older adults (ages 60-85), U. Florida, 2004.

[11] J. A. Bugos, W. M. Perlstein, C. S. McCrae, T. S. Brophy, and P. H. Bedenbaugh, "Individualized piano instruction enhances executive function and working memory in older adults," *Aging & Mental Health,* vol. 11/4 (Jul 2007), pp. 464-471.

[12] J. Ren and X. Luo, "Effect on cognitive processing speed, working memory, and fluid intelligence by learning to play the piano in senior years," *Chinese Journal of Clinical Psychology*, vol. 17/4 (Aug 25, 2009), pp. 396-399.

[13] J. K. Johnson, A. Nápoles, A. Stewart, and W. Max, http://communityofvoices.org/home/meet-the-team/meet-the-team-university-of-california-san-francisco/

The Effect of Visually Presented Lyrics on Song Recall

Yo-Jung Han[1]
[1]Music Education, The Pennsylvania State University, University Park, PA, USA

When learning a song aurally, verbal information (i.e., text) and musical information (e.g., pitches and rhythm) are perceived through the aural channel. However, if a song has much verbal and/or musical information, only using the aural channel might hinder one's ability to learn the song. This study assumed the cognitive load imposed on the aural channel could be reduced if the verbal information is processed in both visual and aural channels by seeing the lyrics. This reduction might provide more capacity to process musical information, leading better recall of the learned song. Therefore, this study aimed to investigate the effect of visually presented lyrics on song recall controlling individual differences in phonological working memory. In addition, the effects of perceived task difficulty, preferred learning modality, and music major were investigated. Members of auditioned choirs were invited to participate. While aurally learning an unfamiliar song, one group saw the lyrics and another group did not. After one individual instructional session, each participant sang the learned song from memory and their singing was recorded. Additionally, participants took a phonological working memory test, a perceived task difficulty report, and a learning style inventory. The recall accuracy of their singing was rated by two experienced music teachers in terms of lyrics, pitches, and rhythm. In the 2 way (Group × Major) MANCOVA with phonological working memory as a covariate, a main effect for major and interaction between group and major were revealed. Seeing lyrics while learning a song was helpful for non-music majors in recalling pitches and rhythm, but not helpful for music majors. For music majors, seeing lyrics might be redundant, which is congruent with the expertise reversal effect. They may process verbal and musical information in a more integrated way than non-music majors.

Music, Autobiographical Memories, and Elderly People

José Davison da Silva Júnior,[*1] Diana Santiago,[#2] Anita Liberalesso Neri[*3]

[*]*Educational Development Department, Federal Institute of Education, Science and Technology of Pernambuco, Brazil*

[#]*School of Music, Federal University of Bahia, Brazil*

[*] *Department of Educational Psychology, State University of Campinas, Brazil*

[1]davisonjr@gmail.com, [2]diana.santiago@pq.cnpq.br, [3]anitalbn@uol.com.br

ABSTRACT

Autobiographical memory can be evoked through sensorial clues, with music being highlighted. Music-evoked autobiographical memories (MEAMs) can be stimulated through popular songs, particularly those which were heard during the 15-24 age range. The elderly evoke more general autobiographical memories. Studies on MEAMs use passive listening to music as the main condition. This study has an additional two conditions. Thus, the participants were involved in three conditions: passive listening to music (condition 1), an activity using clay with background music (condition 2) and composing, performing and audience-listening activities (condition 3). An experiment was carried out with 20 healthy elderly people aged between 65 and 85, using popular Brazilian songs as a stimulus. The autobiographical interview was used to evaluate the content of autobiographical memories. The results showed that both the emission frequency and MEAMs content were higher in condition 3, revealing that direct involvement with music through making music increases the quantity and quality of autobiographical memories.

I. INTRODUCTION

Autobiographical memory covers various cognitive capacities, such as remembering personal events which involve a place, people, emotions and perception, among others (Levine *et al*., 2002). Autobiographical memory can be evoked by various sensorial indicators, such as visual clues, smells and sounds (Willander, Skström, & Karlsson, 2015). When compared with visual clues, music evokes more autobiographical memories and in a more intensive way (Belfi, Karlan, & Tranel, 2015). Sections of popular music serve as clues to retrieve autobiographical memories evoked by music – MEAMs (Janata, Tomic, & Rakowski, 2007).

MEAMs have all the characteristics of involuntary memories, which are conscious memories of personal events that spontaneously come into one's mind (El Haj, Fasotti, & Allain, 2012). Popular songs heard during the 15-24 age range are best remembered and have a stronger relation with autobiographical memories when compared to music in other phases of life (Platz *et al*., 2015).

The elderly prefer and have strong emotional responses to popular music from their youth, when compared with popular music in later periods (Schulkind, Hennis, & Rubin, 1999). When retrieving autobiographical memories, the elderly tend to recapture more general autobiographical information or, that is, information which is not restricted to a single event, known as the overgenerality effect (Ford, Rubin, & Giovanello, 2014).

Research on MEAMs uses passive listening to music as a condition for a relationship with music. This study was extended to include an additional two conditions: an activity using clay with background music and composing, performing and audience-listening activities. The hypothesis was that making music through composing, performing and audience-listening activities will increase the content of autobiographical memories in healthy elderly people.

II. METHOD

A. Participants

20 elderly people (8 men and 12 women, aged between 65 and 85, M=73.9, SD=6.1) who were selected from a group of elderly people in a religious community in the city of Curitiba/Paraná/Brazil. The majority of the elderly people did not work, had completed secondary or higher education and were not professional musicians.

B. Stimulus

The musical stimulus comprised a sample of popular Brazilian songs from the period 1945-1965 which were highly successful when the participants were aged between 15 and 24. The repertoire was selected with the assistance of specific literature from the area (Severiano & Mello, 2006a; Severiano & Mello, 2006b). The list of songs was organized in two groups. The first group comprised ten songs from the period 1945-1954 for the elderly people who were aged between 75 and 85. The second group comprised ten songs from the period 1955-1965 for the elderly people who were aged between 65 and 74.

C. Procedure

The experiment took place during June and July 2015. Each subject took part individually in a single meeting which lasted approximately one hour. The study was developed in three conditions: passive listening to music (condition 1); activity using clay with background music (condition 2) and composing, performing and audience-listening activities (condition 3).

The participants were divided into two groups according to the order that the conditions were presented. Group 1 took part with the conditions in the following order: condition 1, condition 2 and condition 3. Group 2 took part with the conditions in the following order: condition 1, condition 3 and condition 2.

Following signature of the Informed Consent Form, the participant answered a social and economic questionnaire and was submitted to a Mini Mental State Examination (MMSE) to verify if their cognitive functions were preserved.

Then the participants selected three popular Brazilian songs from a list of ten, indicating his order of preference. Soon after this, they were given a Geriatric Depression Scale (GDS) assessment to ascertain their mood. The participants heard the first song selected in the "passive listening to music" condition. An autobiographical interview was then held. The second song selected by the participants was used in the "activity with clay with background music" activity or "composing, performing and audience-listening activities", depending if the participants were from Group 1 or Group 2.

Following participation in the second condition, the second autobiographical interview and distracter activity were carried out. The third song indicated by the participants was used in the "activity with clay and background music" or "composing, performing and audience-listening activities", depending if the participants were from Group 1 or Group 2. The third interview was carried out after the third condition, and Positive and Negative Affect Schedule (PANAS) was applied to verify their mood.

D. Autobiographical Interview

The autobiographical interview (Levine *et al*., 2002) quantifies the content of autobiographical memories. There are previously defined categories which are called details, which can be internal, if they are directly related to the main event, or external if they were more general memories which are not related to the main event.

Internal details may be an event (unfolding of the story, including who was there, reactions/emotions in others, time, one's clothing, and the physical state and actions of others), place (country, city, street, apartment, rooms and location within a room), time (year, season, month, date, day of the week, time of day or clock time), perception (auditory, olfactory, tactile/pain, taste, visual, spatial-temporal) and emotion/thought (feeling states, thought, opinion, expectation or belief).

The external details are semantic (general knowledge or facts), repetitions (do not add new information) and other details (metacognitive statements, editorializing and inferences). The scoring for the detail categories is awarded according to the emission frequency and wealth of content. In addition to the detail categories, the Autobiographical Memory Interview - AMI (specific memory in time and place), time integration (memory of the event within a wider time scale) and episodic richness (impression of experiencing the event again) categories were also included.

Scoring is only attributed for these categories if there is a specific event. The autobiographical interview is held in two parts: a general and specific investigation. In the general investigation, the interviewer invites the participant to recount an autobiographical memory. He then asks if he wishes to say anything else. In the specific investigation, verbal clues are given in order to add further details to the participant's memory.

Only the specific investigation of the autobiographical interview was used to preserve the involuntary nature of the MEAMs (El Haj, Fasotti, & Allain, 2012) and also on account of the overgenerality effect (Ford, Rubin, & Giovanello, 2014).

E. Data analysis

The statistical comparisons between Groups 1 and 2 were carried out according to the logic of intra- and inter-group comparisons. A normality distribution exam was performed, using the Kolmogorov- Smirnov test. As they were anormal, it was decided to adopt nonparametric tests: chi-squared, to compare categorical data; Mann-Whitney for the comparisons between Groups 1 and 2 regarding the scores obtained and Friedman for comparisons between the distributions of each group in the three measurements used (dependent samples).

Firstly, the quality and quantity of the emissions from the elderly people in each of the groups in the three measurement situations were investigated. Then the groups were compared regarding the cognitive evaluation data (MMSE) and two affectivity evaluations (GDS and PANAS). The Statistical Analysis System (SAS) for Windows program, version 9.2., SAS Institute Inc, 2002-2008, Cary, NC, USA was used for the statistical analysis.

III. RESULTS

The emission of details related to internal and external events related by the two groups in the three situations to evaluate autobiographical memories (condition 1 – passive listening to music; condition 2 - an activity with clay and background music and condition 3 – composing, performing and audience-listening activities for Group 1; condition 1 – passive listening to music; condition 3 – composing, performing and audience-listening activities and condition 2 - an activity with clay and background music for Group 2 was counted.

The emission frequency was calculated as per the categories and sub-categories established (Levine *et al*., 2002). The "Other" category was excluded because it included emissions which did not represent autobiographical memories and thus were not of interest to this study. The One Sample Kolmorov-Smirnov test, applied to this data, revealed that the response frequency for condition 3 was statistically different from those observed in conditions 1 and 2, whereas the frequencies observed in conditions 1 and 2 did not differ in a statistically significant manner ($p < 0.001$).

Intra-group comparisons were performed in relation to the performances observed in the three conditions to evaluate the contents of the autobiographical interviews. In Group 1, a statistically significant increase was observed in the scores for the AMI category in condition 2 to condition 3; from the episodic richness category in condition 1 to condition 3 and the total evaluation; from the total evaluation between conditions 1 and 3.

In Group 2, there was an improvement in the scoring from condition 1 to condition 3 in the total emotion/thought, internal emotion/internal thought, AMI, episodic richness and total evaluation categories. Similarly, an improvement in the scoring in these categories between conditions 2 and 3 was also noted. However, there was a reduction in the score in the integrated time category between conditions 1 and 2, but an increased score between conditions 2 and 3.

IV. DISCUSSION

Condition 3 "composing, performing and audience-listening activities" had the potential to evoke a higher number of autobiographical memories than condition 1 "passive listening to music" and condition 2 "activity with clay and background music". This demonstrates the importance of direct involvement with music and not only passive listening to music to increase autobiographical memories. It also shows that even the execution of another creative artistic activity, such as the activity with clay and background music, did not evoke more autobiographical memories than making music. Composing, performing and audience-listening activities are the fundamental processes of music and together with the fundamental possibilities of direct involvement with music, represent the basic categories of musical behavior (França & Swanwick, 2002).

In Group 1, the increased AMI classification following condition 3 when compared with condition 2 showed that the participants detailed more specific autobiographical memories in time and place and not more general memories. Although the elderly people recounted more general memories, direct involvement with music through composing, performing and audience-listening activities favoured evoking more specific memories due to the fact that making music was related to identity and emotion (Creech *et al*., 2013; Dabback & Smith, 2012; Hallam, 2010). Identity and emotion are also related to autobiographical memory, which is fundamental for the self, emotions and personal experiences (Conway & Pleydell-Pearce, 2000). In Group 2, in addition to the increased AMI classification in condition 3 when compared with condition 2, there was also an increase in condition 3 when compared to condition 1, which demonstrated that the content of the autobiographical memories remained more general in condition 1. It was concluded that there is greater specificity of autobiographical memory in time and place when the subjects are involved in composing, performing and audience-listening activities when compared to passive listening to music and an activity with clay and background music.

In Group 1, the episodic richness was higher after condition 3, when compared to condition 1. Evoking more general memories after passive listening to music did not bring a feeling of experiencing the past, as occurred following composing, performing and audience-listening activities. In Group 2, the episodic richness was greater in condition 3, when compared to the other categories or, that is, participation in condition 3 before condition 2 evoked a greater sensation of going back to a specific moment in time and place from the centrality of making music. It is observed that there is a greater intensity of evoking autobiographical memory when the person is involved in making music, when compared to passive listening to music and an activity with clay and background music. Direct contact with music makes the experience of evoking memories more intensive.

The order of participating in conditions 2 and 3 (Group 2) showed differences in the content of the total emotion/thought, internal emotion/thought and integrated time categories. The participants evoked a richer emotional content in condition 3 than from conditions 1 and 2. In addition, a stronger integration of remembering an episodic event within a wider timescale occurred with the Group 2 participants or, that is, there was a more detailed description of contextual information which took place before or after the recollected event. That fact that these results only took place in Group 2 demonstrates that the Group 1 participants did not have increased emotional content or integrated time in condition 3 because they were previously involved with condition 2. It can be inferred that the other artistic activity carried out before composing, performing and audience-listening activities could minimize the potential of making music evoking a higher emotional and integrated time content.

In Groups 1 and 2, the total evaluation of the content of the autobiographical memories was higher following condition 3 than conditions 1 and 2. The sum total of the scoring attributed by the content of the details of place, time, perception and emotion/thought, integrated time, episodic richness and episodic repetition (total score) expresses the wealth of content of autobiographical memories, which was greater following composing, performing and audience-listening activities.

These results, together with those of a higher frequency of emissions of autobiographical memories following condition 3, when compared to conditions 1 and 2 in both groups, prove our hypothesis that direct involvement with music through composing, performing and audience-listening activities increase the content of autobiographical memories. Autobiographical memories increased both in quantity (emission frequency) and quality (memory content) following direct involvement with music through making music.

ACKNOWLEDGMENT

This study was conducted as part of the first author's Doctorate program in Music, held at the School of Music of the Federal University of Bahia (UFBA), Brazil, and with support from the National Council for Scientific and Technological Development (CNPq) – Brazil.

REFERENCES

Belfi, A. M., Karlan, B., & Tranel, D. (2015). Music evokes vivid autobiographical memories. *Memory*, 10, 1-11.

Conway, M. A., & Pleydell-Pearce, C.W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological Review*, 107 (2), 261-288.

Creech, A., Hallam, S., Gaunt, H., Pincas, A., McQueen, H., & Varvarigou, M. (2013). The power of music in the lives of older adults. *Research Studies in Music Education*, 35, 1, 83-98.

Dabback, W. M., & Smith, D. S. (2012). Elders and music: empowering learning, valuing life experience, and considering the needs of aging adult learners. In: McPherson, G. E., & Welch, G. F. (Ed.). *The Oxford Handbook of Music Education*, vol. II. New York, Oxford University Press, 229-242.

El Haj, M., Fasotti, L., & Allain, P. (2012). The involuntary nature of music-evoked autobiographical memories in Alzheimer's disease. *Consciousness and cognition*., v. 21, 238-246.

Ford, J. H., Rubin, D. C., & Giovanello, K. S. (2014). Effects of task instruction on autobiographical memory specificity in Young and older adults. *Memory*, vol. 22, n. 6.

França, C. C., & Swanwick, K. (2002). Composing, audience-listening and performance in music education: theory, research and practice. *Em Pauta*, v.13, n.21, 5-41.

Hallam, S. (2010). Music education: the role of affect. In: Juslin, P. N., & Sloboda, J. *Music and emotion: theory, research, applications*. New York: Oxford University Press, 791-817.

Janata, P., Tomic, S. T., & Rakowski, S. K. (2007). Characterisation of music-evoked autobiographical memories. *Memory*, 15, 8, 845-860.

Levine, B., Svoboda, E., Hay, J. F., & Winocur, G. (2002). Aging and autobiographical memory: Dissociating episodic from semantic retrieval. *Psychology and Aging*, 17 (4), 677-689.

Platz, F., Kopiez, R., Hasselhorn, J., & Wolf, A. (2015). The impact of song-specif age and affective qualities of popular songs on music-evoked autobiographical memories (MEAMs). *Musicae scientiae. The jornal of the european society for the cognitive sciences of music*, v. 19, n. 4, 327-349.

Schulkind, M. D., Hennis, L. K., & Rubin, D. C. (1999). Music, emotion, and autobiographical memory: They're playing your song. *Memory & Cognition*, 27 (6), 948-955.

Severiano, J., & Mello, Z. H. (2006a). *A canção no tempo: 85 anos de músicas brasileiras (vol.1: 1901-1957)*. 6ª ed. São Paulo: Ed. 34.

Severiano, J., Mello, Z. H. (2006b). *A canção no tempo: 85 anos de músicas brasileiras (vol.2: 1958-1985)*. 5ª ed. São Paulo: Ed. 34.

Willander, J., Sikström, S., & Karlsson, K. (2015). Multimodal retrieval of autobiographical memories: sensory information contributes differently to the recollection of events. *Frontiers in Psychology*, v. 6.

Enhancing working memory for actions through entrainment to an auditory cue

Kathleen Wyeth[1,2], Kate Stevens[1,2] and Jia Hoong Ong[1]
[1]MARCS Institute for Brain, Behavior & Development, Western Sydney University, Sydney, Australia
[2]School of Social Sciences & Psychology, Western Sydney University, Sydney, Australia

Studies of verbal memory suggest that the presence of a regular auditory beat aids learning and memory. We investigate facilitation from an auditory beat for working memory for actions. On each trial of a recognition task, four whole body actions were presented as a study set, followed immediately by a test set of four actions. The task was to recognize whether the test set actions were the same or different from the study set. Half of the trials were different with a single action replaced from a set of eight actions. The presence of an auditory beat coinciding in time with presentation of the visual action stimulus was manipulated across study and test sets. The crossing of the auditory beat being either present or absent across study and test produced four conditions: beat study-beat test; silent study-silent test; beat study-silent test; silent study-beat test. A fifth condition was included, to decouple visual and auditory entrainment, wherein the visual action and auditory beat were mismatched in time. It was hypothesized that i) recognition is greater when an auditory cue is synchronized with the action items during study than when an auditory cue is mismatched or absent; and ii) relatively high accuracy is elicited when study and test conditions match ("encoding specificity"). Thirty participants completed 100 randomised trials. The first hypothesis was supported with significantly greater accuracy recorded in the beat study-beat test condition (d'=1.15) than in the beat study-silent test condition (d'=0.95). Poorest accuracy, as hypothesized, was recorded in the mismatched condition (d'=0.83). Relatively good performance when silent study and test conditions matched, supported the hypothesis (d'=0.94). Facilitation in accuracy did not manifest in reaction time; there was no speed-accuracy trade-off. Results are interpreted using Dynamic Attending Theory: the regular auditory cue heightens arousal and accuracy whereas response timing is locked to the cue.

How do you remember all the words? Singing along facilitates memory for song lyrics

Lucy M. McGarry[1], Adrian M. Owen[1], Jessica A. Grahn[1]
[1]The Brain and Mind Institute, Department of Psychology, University of Western Ontario, Canada

Background
Many individuals have an exceptional capacity to remember the words to numerous songs from the past. Several factors may enhance memory for song lyrics versus memory for other spoken items such as speeches or poetry; these include repeated listening, liking, and the presence of melody to assist with encoding of words. Individuals also tend to interact with music by singing or dancing in synchrony, and this could add an additional source of motoric encoding.

Aims
Here we examine whether singing during encoding enhances subsequent lyric memory; specifically, whether singing along vs. speaking along or simply listening quietly to song facilitates memory for song lyrics. We also examined context (a mixture of sung/listened trials or spoken/listened trials versus blocks of sung-only, spoken-only, or listened-only trials).

Methods
Sung and spoken sentences were presented audio-visually. Participants produced or listened to each sentence in a ‘mixed’ list (produced and listened) or ‘blocked’ list (produced only or listened only), followed by free recall tests.

Results
We found significant ‘production effects’ for song and speech. As in prior research, this difference occurred only in a mixed list context, implicating distinctiveness as a factor in the production effect.

Conclusions
Singing along to music improved word recall, as did speaking. Enhanced long-term memory for lyrics may partially be due to the tendency to sing in synchrony with the piece. It is likely people sing along with song more often than they recite along with speech in everyday life, which helps to explain findings of enhanced long-term memory for lyrics versus speech even though both have the capability to enhance memory.

Investigating autobiographical memory recall in dementia and healthy older adults following exposure to familiar popular music

Renee Schapiro[1,2], Denise Evert[1] and Daniel Müllensiefen[2]
[1]Psychology Department and Neuroscience Program, Skidmore College, Saratoga Springs, NY, USA
[2]Psychology Department, Goldsmiths, University of London, London, UK

Autobiographical memory (ABM) greatly contributes to an individual's sense of identity, yet significantly deteriorates as dementia progresses. Previous research has found that music facilitates ABM recall in patients with dementia. While past studies often used non-lyrical classical music (e.g., Vivaldi's Four Seasons: Foster & Valentine, 2001; García et al., 2012; Irish et al., 2006) and/or self-selected music (El Haj et al., 2012), the present study utilized familiar popular music from two time periods in participants' lives. The independent variables included the life period of musical selections (between-subjects: adolescence, adulthood), the auditory condition (within-subjects: happy song, sad song, silence) and the health of participants (25 healthy older adults, 14 dementia patients). The dependent measures included changes in emotional responses before and after exposure to the auditory condition (as measured by the Positive and Negative Affect Schedule: PANAS; Watson & Clark, 1999) and scores on an autobiographical memory interview (adapted from Foster & Valentine, 2001 & García et al., 2012). Participants completed three sessions, each spaced approximately one week apart. PANAS results demonstrated the successful induction of positive or negative affect following happy or sad songs, respectively, both in dementia and healthy participants. The change in positive affect was strongest for music from their adolescence. Contrary to previous research, music did not significantly benefit ABM recall as compared to silence for either participant group. It is possible, therefore, that familiarity with music may not be the factor driving memory recall improvements found in other studies; rather, the effect may be due to arousal levels induced by music and/or the personal significance of the music. Proposed future directions will be discussed.

Differential use of cognitive strategies in a musical transcription task: cues from implemental memory

Maria Teresa Moreno Sala[1,2], Ruth Cruz de Menezes[1,2], Matthieu J. Guitton[3,4]
[1]Laboratoire de recherche en formation auditive et didactique instrumentale (LaRFADI), Faculty of Music, Laval University, Quebec City, Canada
[2]Observatoire International de création et recherche en musique (OICRM), Canada
[3]Faculty of Medicine, Laval University, Quebec City, QC, Canada
[4]Institut Universitaire en Santé Mentale de Québec, Quebec City, Canada

Transcribing musical dictation involves many cognitive processes: perception, spatio-temporal organization, memory short-term storage, analysis and transformation of auditory percepts into written signs. However, the exact cognitive strategies and the processes underlining their use to solve this task are still unknown. The purpose of the present study is to categorize the types of strategies used to transcribe a melodic dictation, observe their frequency of utilization, identify the most efficient strategies, and verify whether mnemonic capacities were linked to the use of the strategies. For this study, 38 undergraduate musical students were asked to complete a melodic musical dictation and write simultaneously the strategies they used. Then, the subjects were exposed to an object recognition task to assess their visual short-term memory span. As results, two complementary types of cognitive strategies emerged from the analyses of the student's descriptions and were classified as non-tonal strategies (not linked to tonality cues) and tonal strategies (relying on tonal associative cues). Several statistical analyses were done. Mainly, the global efficacy of tonal strategies was significantly better than for non-tonal strategies. In general, subjects who performed better used more strategies, more often, and more effectively. Furthermore, the efficacy of non-tonal strategies correlated with memory span in some categories of subjects (ρ=0.750, p<0.05). Our results suggest a differential use of non-tonal and tonal strategies in order to complete a musical transcription task and the existence of an underlying relationship between mnemonic capacities and the optimal use of some cognitive strategies. These results may contribute to a better understanding of the cognitive processes used to transcribe melodic dictation for improving teaching strategies as well as aural skills methods.

The Effects of the Instrument Played on the Working Memory Capacity of Musicians

Stephanie Whittaker[1] and Margaret Cousins[1]
[1]Psychology Department, The University of Chester, Chester, Cheshire, UK

Previous studies have shown differences between musicians and non-musicians in visuo-spatial memory, verbal memory and auditory working memory. The research presented aimed to explore possible differences between musicians, delineated by the multi-tasking requirements of the instrument they play. Participants were musicians (N=32) divided equally into four categories: singers, brass players, woodwind/string players and finally organists, percussionists and conductors. Participants were asked to complete a questionnaire and undertake three memory tests: digit-span, Corsi block and a newly designed music-span test in both forwards and backwards conditions. The music-span consisted of atonal note rows selected from a totally chromatic sequence of notes with no reference to a diatonic framework. These increased in length following the format of the digit-span test. The results indicated that singers were strongest across the forwards tests. Multi-tasking musicians (organists, percussionists and conductors) performed best in the forwards and backwards digit-spans tasks, but not the spatial test.

Better late than never (or early): Late music lessons confer advantages in decision-making

Kirsten E. Smayda[1], W. Todd Maddox[1], Darrell A. Worthy[2], Bharath Chandrasekaran[1,3]
[1]Psychology Department, The University of Texas at Austin, Austin, TX, USA
[2]Psychology Department, Texas A & M, College Station, TX, USA
[3]Communication Sciences and Disorders, The University of Texas at Austin, Austin, TX, USA

Music training early in childhood, relative to later, confers cross-domain sensory and motor advantages. Both the sensory and motor systems mature early in life, and training during this critical period in development, relative to later training, is suggested to confer a greater benefit on the sensory and motor systems of the brain that persist past childhood. Here we ask whether late music training enhances decisional processes, given that brain regions that mediate cognitive and decisional processes show a more protracted maturation into adolescence. We recruited young adult musicians who began music training before the age of eight (early-trained; ET), after the age of eight (late-trained; LT) and non-musicians (NM) to perform a classic decision-making task, the Iowa Gambling Task. We found that late-trained musicians showed an advantage in decision-making relative to both early-trained musicians and non-musicians. To explore possible mechanisms of the late-musician performance advantage, we conducted a number of individual differences analyses between and within groups. Results suggest no group differences in measures of working memory, inhibition, and short-term memory. Lastly, we found that working memory measures were predictive of performance in the Iowa Gambling Task in only the early musician group, suggesting differential mechanisms for decision-making depending on when an individual began playing music. These results argue for a potential benefit of music instruction later in childhood on critical decisional processes.

The Benefits of Music Making on Cognition in Older Adults

LinChiat Chang[1], Anna M. Nápoles[2], Alexander Smith[3] and Julene K. Johnson[4]
[1]Independent Consultant, San Francisco, CA, USA
[2]Department of Medicine, University of California, San Francisco, CA, USA
[3]UCSF and SF VAMC Geriatrics and Palliative Care, San Francisco, CA, USA
[4]Institute for Health & Aging, University of California, San Francisco, CA, USA

Background
Research has demonstrated that older adults who sing or play music have enhanced cognitive functioning, but most studies focused on persons of higher socioeconomic status.

Aims
This study examines the link between music and cognition in a diverse sample of older adults and whether propounded benefits vary by socioeconomic status.

Methods
Data was extracted from the 2014 Health and Retirement Study (HRS), designed to represent the general population of Americans over 50 (mean age=68). Analyses compared older adults who sing or play music with those who do not (n=1,496) on verbal memory (immediate and delayed word recall), executive function (serial 7's, numeracy, backwards count), and semantic memory (object, person naming). Multivariate regression models tested whether potential benefits of music on cognition vary by socioeconomic attributes including education, income and assets, race/ethnicity, sex, and age. Sample weights were applied to correct for nonresponse bias, design effects, panel attrition and mortality.

Results
Older adults who sing or play music performed better than those who do not make music on tests of verbal memory ($p < 0.001$), but the two groups did not differ in performance on executive function or semantic memory tests. The positive effect of music making on verbal memory remained robust ($p < 0.001$) even after controlling for socioeconomic variables. A significant interaction with education ($p < 0.05$) implied stronger effects of music making on verbal memory among older adults with more years of education than those with fewer years. A marginal interaction with age ($p < 0.10$) suggested that potential benefits of music making on verbal memory may weaken with advancing age. Effects of music on cognition did not vary by income and assets, race/ethnicity, or sex.

Conclusion
Results suggest that music making is related to superior verbal memory among older adults, particularly among those with more years of education.

Outcomes of participation in a piano-based recreational music making group for adults over 50: A retrospective study

Olivia Swedberg Yinger[1], Vicki McVay[1], Lori Gooding[2]
[1]School of Music, University of Kentucky, Lexington, KY, USA
[2]College of Music, Florida State University, Tallahassee, FL, USA

Musical participation has many benefits, including improved health, reduced stress, and social connectedness, that have been linked to successful aging. The purpose of this study was to investigate perceptions of a 10-week piano-based music and wellness group program designed for adults age 50 and older. Specifically, researchers examined which program aspects participants perceived were beneficial, needed improvement, and impacted their quality of life and emotional well-being. This study was a retrospective review of data gathered from participants ($N = 26$) who took part in at least one 10-week piano-based music and wellness group. Measures included a feedback form and quality of life questionnaire. Responses to Likert-type scale questions were analyzed using descriptive statistics. Written comments were analyzed using Linguistic Inquiry and Word Count (LIWC) software and content analysis. Results indicated that participants enjoyed the group learning environment, gained both cognitive and stress reduction benefits, and were highly likely to continue to play the piano. Suggested improvements included increasing class offerings, increasing access to pianos for practice, and providing email/social media notifications. Participants reported high levels of quality of life, particularly as it relates to independence, health, and support from family, friends, or neighbors. Linguistic analysis of comments revealed higher percentages of positive emotion words, social words, and leisure words compared to negative emotion words. A content analysis of written comments revealed reasons for joining the program, including stress relief, self-satisfaction, cognitive stimulation, love of music, mental health benefits, personal enjoyment, and coping skills. This study provides information about a recreational music-making program specifically designed for older adults which can be used to develop music programs that encourage lifelong musical participation and successful aging.

Does Aging affect the Voice of Trained Professional Singers?

Thomas Akshintala, Rashmi Pai, G.S.Patil

Abstract— Music is the art of arranging sounds in time so as to produce a continuous, unified and evocative composition through melody, harmony and rhythm. The music of India includes multiple varieties of folk music, pop and classical music. Indian classical music is one of the oldest forms of music in the world. It has its roots in divers area such as the ancient religion in Vedic hymns, tribal chants, devotional temple music, and folk music. Indian classical music is a raga system. Vocal aging is a gradual process, with several vocal changes occurring at different stages of an adult's life. With aging, the larynx undergoes anatomical and physiological changes that have an impact on the vocal quality**.** Aim of the study was, to compare the age related changes in intensity, frequency, HNR among trained professional carnatic singers.31 trained female carnatic singers in the age range of 13 to 60 years, divided into 3 groups as; below 20 years, 20-40 years and 40 to 60 years, were considered for the study. Voice recording was carried out in a quiet environment using a digital voice recorder placed at a distance of 6cm away from the singer's mouth. Subjects were asked to sing "Shankarabharanam" The ragas were recited in the form of "alap" (phonating) at the comfortable level used by them.

Parameters considered for study were, mean frequency and intensity, minimum frequency and intensity, maximum frequency and intensity, frequency and intensity range and HNR. Acoustic analysis was carried out using Dr. Speech version 4 software. Paired sample t test was applied to compare the parameters of the various age group using SPSS 17 software.The mean, standard deviation values obtained and compared across three groups. Mean frequency, Minimum frequency and maximum frequency found to be higher among Group II, Mean intensity, maximum intensity, intensity range and HNR higher in Group I. Although literature shows changes in voice with ageing, no significant difference observed among the frequency, intensity and HNR. It was statistically evident from the present study that, ageing has very minimal effect on trained professional singers. No statistically significant difference observed among all the three age groups. This could be attributed to their years of practice, vocal warm ups, habits, vocal hygiene followed and factors alike.

Keywords— Aalap, Acoustic parameters of voice, Carnatic music, ageing, professional singers.

T. Akshintala is with Institute of Speech and Audiology as a Clinical Supervisor; Sweekaar Academy of Rehabilitation Sciences; Secundrabad; Osmania University INDIA(Phone:+91 9491353712;E-mail:akshintalathomas@gmail.com)

R. Pai , is an Associate Professor at Institute of Speech and Audiology; Sweekaar Academy of Rehabilitation Sciences; Secundrabad; Osmania University INDIA(E-mail: dessairashmi@gmail.com)

G.S.Patil is with Ali Yaver Jung National Institute for Hearing Handicapped, SRC, Secunderabad. Osmania University INDIA(E-mail: gourishanker07@rediffmail.com)

I. Introduction

Voice can defined as laryngeal modification of pulmonary air stream and requires systemic coordination between the respiratory, phonatory,resonatory and articulatory apparatus. Aging is a complex process of biological events that changes the structure and function of varies parts of body. Vocal aging is a gradual process, with several vocal changes occurring at different stages of an adult's life. With aging, the larynx undergoes anatomical and physiological changes that have an impact on the vocal quality (Linville, 2001).

E.T.Stathoulas,J.E.Huber and J.E.Sussman (2001) Carried out a study investigating changes in acoustic characteristics of voice across life span among 192 individuals. Fundamental frequency (F0) and Signal to Noise ratio (SNR) showed gender difference across the life span unlike SPL. Variability of F0, Sound pressure level (SPL) and SNR followed a nonlinear trend which was found to be higher in younger than older age group.

E.D'haseleer,et.al., in their study, measured and described the effects of aging on middle aged premenopausal women. Both subjective and objective tests of the vocal assessment were used. Results of the study showed reduced fundamental frequency, more restricted vocal frequency range and lower minimum fundamental frequency, reduced intensity range, reduced maximum intensity levels among middle aged premenopausal women when compared to young women

Singing is product of delicate balancing of physiological control, artistry, and technique (Teachey, kahane & Beckford, 1991). The uniqueness of Indian classical music is in its melodic structure, which is based involved in voice production on the system of "Raga" system.

Indian classical musical system may be broadly classified into Carnatic and Hindustani systems. The scale of carnatic music system comprises seven tones or swaras – sa, ri, ga, ma, pa,da,and ni. Such a scale is known as 'sampurna raga'. Indian classical music is predominantly improvisatory from tonal music

Effects of training can be speculated on the process of vocal aging. Brown et. al (1991) in their study, suggested that professional singers are more resistant to effects of aging as they master the techniques to preserve optimal laryngeal functioning with aging and and also practice a good vocal hygiene. A study by D.Hazelett and M.J.Ball(1996) compared acoustic characteristics among two trained singers with varying years of experience. The results showed lower F0 and lower jitter in subject, with higher and history of smoking. Virtually identical HNR scores were observed though subject 2 with lesser experience and no history of smoking showed greater hoarseness as compared to subject 1. VOT indicated no significance difference among the speakers.

(Koufman et. al(1996); Sonnien & Hurme (1998) in their study, concluded that the singers exhibit less muscular tension than the non singers, while singing. They also suggested that the singers tend to sound younger as they advance in age, than their non professional counter parts due to their usage of voice conservation stratergies.

Brown et al 1991,1993.,) suggested that professional singers do not follow the patterns of aging change observed in non singers and are somewhat immune from the effects of ageing. Professional singers tend to maintain stable fundamental frequency (Fo) levels throughout their adult life in contrast to the non singers and also do not follow the patterns of ageing as seen in others.

II. Need For the study

Literature of review shows evident changes in voice of a normal individual. But there is dearth of literature on effect of ageing in trained professional singers. Hence the present study aimed at comparing the age related changes in frequency, intensity and harmonic to noise ratio parameters among professional carnatic singers.

Aim of the study, was to profile and compare the age related changes in intensity, frequency, harmonics to noise ratio among trained professional carnatic singers.

III. Method:

Present study was carried out at Sweekaar Academy of rehabilitation science; Secunderabad Telangana,to compare the acoustic parameters of voice across age groups.

31 trained female carnatic singers in the age range of 13 to 60 years were considered for the study. Subjects were divided into 3 groups as; below 20 years, 20-40 years and 40 to 60 years.

Inclusion Criteria:

- Trained professional Carnatic Singers
- Minimum of 4 years formal vocal training.
- Singing practice of at least 20 hours per week

Exclusion criteria:

- No history of speech, language, or hearing problems
- No history of alcohol or tobacco intake,
- No History of smoking,
- No history of laryngeal trauma.

Voice recording was carried out in a quiet environment using a digital voice recorder placed at a distance of 6cm away from the singer's mouth. Subjects were asked to sing "Raga Dheer Shankarabharanam" The written version of musical notes in ascending order was provided to the participants. The ragas were recited in the form of "alap" (phonating) at the comfortable level generally used by the singer.Recordings of the same, were transferred to the computer system with the voice analysis software, Dr.Speech installed in it. Dr. Speech is a comprehensive speech/voice assessment software system, that is intended for use with professionals in voice and speech fields. Parameters considered for study were, mean frequency and intensity, minimum frequency and intensity, maximum frequency and intensity, frequency and intensity range and harmonic to noise ratio (HNR). Acoustic analysis of the obtained samples was carried out . Multiple analysis of variance (SPSS 17.0) was applied to compare the parameters of the various age groups considered in the study.

Null Hypothesis: There will be significant difference in voice parameters between the three groups.

IV. Results And Discussion

The present study compared the fundamental frequency (F0), frequency range and harmonics to noise ratio (HNR) mean intensity, minimum and maximum intensity, and intensity range among professional carnatic singers. The mean and the standard deviation values obtained were obtained and compared across the three groups.

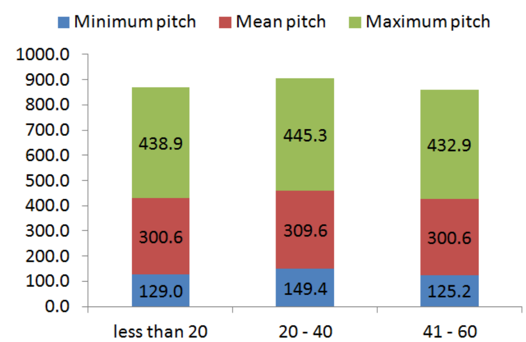


Figure.1.Indicating pitch of three groups

Mean frequency, Minimum frequency and maximum frequency were found to be higher among Group II participants in the age range of 20 to 40 years. Whereas frequency range and Harmonic to noise ratio was found to be higher in group I that comprised of participants with a younger age as shown in figure 3. Statistically no significant difference was observed across the groups for all the frequency parameters as well as harmonic to noise ratio.

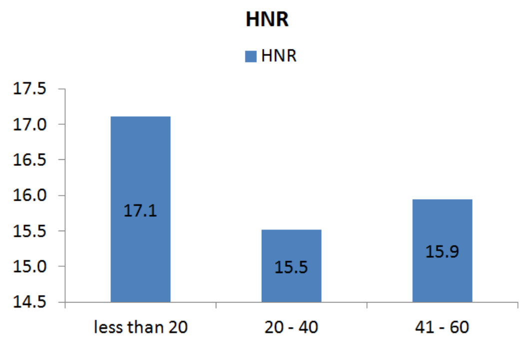


Figure.2.Indicating Harmonics to noise ratio of three groups

Mean intensity, maximum intensity as well as intensity range was found to be higher in Group I as compared to the older groups. Minimum intensity was found to be higher among the Group II participants which was similar to the finding seen for minimum frequency. Intensity parameters also showed a similar trend of on MANOVA test with no statistically significant difference across the age groups as shown in Table 1 below.

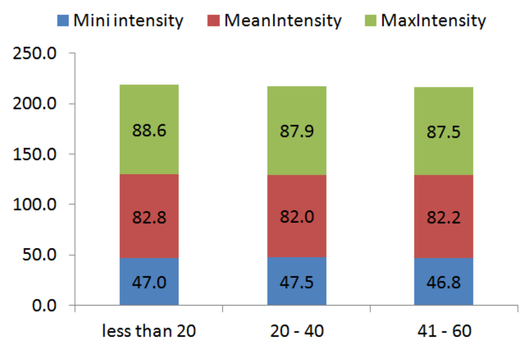


Figure.3.Indicating the Intensity of three groups

Although literature shows evident changes in voice with ageing, there was no significant difference observed among the frequency and intensity parameters as well as harmonic to noise ratio among these trained professional voice users. This study is in agreement with the earliear studies like that by Brown et. Al. (1991) and (Koufman et. al(1996); Sonnien & Hurme (1998) , which also suggested that the professional voice users/ singers are more immune to the effects of ageing. These findings can be attributed to the professional training received by the singers which involves regular vocal warm ups and exercises and a good vocal hygiene program followed and carried out in a systematic manner.

In conclusion, it was statistically evident from the present study that, ageing has very minimal effect on the voice of a trained professional voice user. No statistically significant difference was observed among the age groups of less than 20 years, 20 to 40 years and 40 to 60 years.

Table 1 showing the statistical data.

	Dependent Variable	df	Mean Square	F	Sig
GROUPS	MINPITCH	3722.482	1861.241	.897	.420
	MEANPITCH	899.170	449.585	1.722	.198
	MAXPITCH	1209.365	604.683	1.525	.236
	RANGE	1210.728	605.364	.292	.749
	HNR	15.958	7.979	.907	.416
	MININTENSITY	.790	.395	.063	.939
	MAXINTENSITY	13.531	6.766	.721	.495
	MEANINTENSITY	24.221	12.111	.931	.407
	RANG	9.486	4.743	.340	.714

This could be attributed to their years of practice, vocal warm ups and habits, vocal hygiene followed and factors alike. Hence, there is a need for an extensive study of a similar kind in future with a larger sample size and taking into consideration all the related factors.

A. References

[1] Banerjee, B.M., Sengupta, R., Sengupta, S., Nag, D., Dutta, A.K. Ganguli, N.R.,(1983). Spectral Characteristics of vowels Sung in Hindustani Sastriya Sangeet. *Journal of the Acoustic Society of India, 11(4),*45-47.

[2] Boon , D. (1977) *The voice and voice therapy*. Englewood cliffs, New Jersey Prentice Hall.

[3] Bartholomew, W.Y. (1934). A physical definition of “good voice” quality in the male voice. *The Journal of Acoustic Society of America, 79,* 2028-2033.

[4] Bunch, M. (1982). *Dynamics of singing voice.* New York: Springer, Verlag.

[5] Burns, P., and Austin (1986) Acoustic Analysis of the Underlying Voice Difference Between two Groups of Professional Singers Opera and Country and Western.*Laryngoscope,96,* 549-559:

[6] SpringeCarrol, C. (1974). Longevity of Vocal. *Folia Phoniatrica.* 26, 293-294.

[7] Case, J.L.(1991). Clinical management of Voice disorders. Austin pro-Ed.

[8] Cahayadevi,. (2007). *Singers Formants in Indian Classical singers- Carnatic V/s Hindustan.* Student researcher at AIISH: Articles based on Dissertation, Vol IV

[9] Danielous, A. (1968) . *The Raga of North Indian Music.* London: Barrie and Rockliff, The Cresset Press.

[10] Datta, A. K., & Sengupta, R., (2001). Studies on identification of raga using short pieces of taan: A Signal processing approach. In proceeding of the *6th International Workshop on Recent Trends in Speech and Music and Allied Signal Processing.* New Delhi.

[11] D’haeseleer E, Depypere H, Claeys S, Van Borsel J, Van Lierde.,KM., *Vocal Characteristics of Middle-Aged Premenopausal Women*., Journal of Voice, Vol. 25, No. 3, pp. 360-366

[12] Durga, S. A. K. (1997). Voice culture with special reference to south Indian music. Indian Musicology Society, Baroda, India.

[13] E.T.Stathoulas.,J.E.Huber & J.E.Sussman (2001). *changes in Acoustic characteristics of voice across the life span measures from individuals 4 to 93 years age*., JSLHR., volume 54, 1011 to 1021, August 2011.

[14] Greene, M. C. L. (1972). *The voice and its disorders.* (Ed. 3), London: Pittman Medical Publishing Co.

[15] http/www.indianmelody.com

[16] Green & Mathaison (1995). Voice disorders. London: Whurr Publishers

[17] Hirano, M., Kiyokawa, K. and Kurita, S. (1988), "*Laryngeal muscles and glottic shaping*," in Vocal physiology: Voice production mechanisms and functions, edited by O. Fujimara (Raven Press Ltd., New York) 49-65

[18] Jairazbhoy, N. A. (1971). *The raga of North India music: Their Structer and Evaluation.* London: Faber and Faber.

[19] Jiang, J., Lin, E. and Hanson, D. G. (2000), "*Voice disorders and phonosurgery i: Vocal fold physiology,*" Otolaryngologic Clinics of North America. 33, 699-718.

[20] Koufman, J.et.al, 1996 Laryngeal Reflex consensus conference report journal of voice, 10(3), 215 -216.

[21] Krishnaswamy, A., (2003). Application of Pitch tracking to south Indian classical music. In *IEEE International Conference*

[22] *on Acoustics, Speech and Signal processing (*pp.557-560*),* Hong Kong, Chain.

[23] Krumhansl, C. L. (1990). *Cognitive Foundations of Musical Pitch.* Oxford: Oxford University Press.

[24] Kiesgen, Paul. "Breathing." *Journal of Singing* 62, no. 2 (2005): 169-171.

[25] Lawrence, V. L. (1979). Laryngological observation on the belting. *Journal of Research in Singing, 2,* 26-28.

[26] Luchsinger, R. L, & Arnold, G. E. Voice, (1965). *Speech and Language.* Constable and Co. Ltd.,

[27] Linville SE. Aging of the larynx. In: Linville SE, ed. Vocal Aging. Canada:Singular Thomson Learning; 2001:37–53

[28] Linville SE, Fisher HB. *Acoustic characteristics of women's voices with advancing age*. J Gerontol. 1985;40:324–330.

[29] McCare, C.R. & Morris, R. J. (2004). Comparison of Voice Onset Time for Trained Male Singers and Male Nonsingers During Speaking and Singing. *Journal of Voice, 19*, 420-430.

[30] Morris R, Brown W. *Age-related voice measures among adult women*.J Voice. 1987;1:38–43

[31] Mendes, A.P. Rothman, H.B. Sapienza, C. & Brown, W.S. (2003). Effects of Vocal Training on the Acoustic Parameters of the Singing Voice. *Journal of voice.17*, 529-43

[32] Reck, D. B. (2005). India/South India. In J. T. Tinton & L. Fujie (Eds.), *Worlds of Music: An Introduction to the Music of the World's peoples. (pp. 196-231).* Belmont, CA: Thomson Learning Inc.

[33] Ross, J. (1992). Formant Frequency in Estonian Folk Singing. *Journal of Acoustic Society of America,* 91, 3532-3539.

[34] Rammohan, A. (2000). *The Handbook of Tamil culture and heritage.* International Tamil Language Foundation. PA: National Publishing Co.

[35] Miller,(1985). Intraindividual Parameters of the Singers Formant. *Folia Phonation,* 37, 31-35.

[36] Savitri, S. R. (2006). Overview of professional voice care. *Journal of ITC Sangeet Research Academy*, *20.* 258-26

Subcortical Processing during Emotional Encoding of Sensory Dissonance

Fernando Bravo[1,2], Emmanuel A. Stamatakis[3] and Martin Rohrmeier[1]
[1]Dresden Music Cognition Laboratory - Institut für Kunst- und Musikwissenschaft, TU Dresden, Dresden, Germany
[2]Centre for Music and Science – Faculty of Music, University of Cambridge, Cambridge, UK
[3]Division of Anaesthesia – Clinical Neuroscience, University of Cambridge, Cambridge, UK

The emotional evaluation of tonal dissonance seems to be strongly influenced by culture, as demonstrated by studies that have documented its variations across different cultures and its historical transformation through the history of western music. However, it has been proposed that judgments of sensory dissonance could be culturally invariant and largely independent of musical training. Empirical findings further suggest that sensory dissonance could be initially processed at relatively basic auditory levels. Interestingly, it has been proposed that the brainstem reflexes could be involved in emotion induction processes aroused by music, a hypothesis that converges with the assumption of a potential "fast" subcortical route supporting affective responses to music. The present fMRI experiment indirectly assessed this issue, by investigating the effects of sensory consonance/dissonance on the cognitive and neural mechanisms underlying emotion recognition in sound. Results showed that distinct levels of dissonance exerted differential influences on the right angular gyrus, an area implicated in mental state attribution and attention reorienting processes. Importantly, an involvement of the brainstem was evidenced whilst participants judged the strong dissonant sounds (compared with consonant sounds). Specifically, the right inferior colliculus showed stronger functional connectivity with the right angular gyrus, suggesting that early sensory processes in the auditory pathway might be contributing to relatively complex cognitive functions (i.e. judgments of valence assigned to increasing levels of sensory dissonance). Overall, our findings further substantiate previous research, showing an involvement of subcortical processing stages during the emotional encoding of sensory dissonance.

The influence of timbre and expressive intent on musical communication of emotion

Kit Soden[1], Jordana Saks[2], Stephen McAdams[1]
[1]Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT), Schulich School of Music, McGill University, Montreal, QC, Canada
[2]Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT), Cognitive Science Program, McGill University, Montreal, QC, Canada

During a musical performance, does the listener perceive the emotion the performer is expressing? Is there affective information composed into the melodic material of a dramatic work? How is the intended expression influenced or affected by the range of timbral variation available on a given instrument? Our study seeks to understand the communication process involving composer, performer, and listener by abstracting melodic lines from their original context within operatic works, and then reinterpreting these excerpts using different timbres and expressive affects. Three short excerpts from Romantic-era operas were chosen, based on their clarity in signifying one of five basic discrete emotions (anger, fear, happiness, love, sadness). For the reinterpretations, soprano voice and five instruments (flute, English horn, clarinet, trumpet, cello) were chosen to include a wide variety of timbral and expressive characteristics. Six professional musicians performed the excerpts while attempting to express one of each of these five basic emotions during their performance, yielding a total of 90 excerpts. Participants (N=40) were asked to listen to each excerpt and to complete two tasks: 1) during listening they continuously rated the overall emotional intensity from weak to strong using a slider that provided force feedback; 2) at the end of the excerpt they picked from the emotion (s) those that they perceived as being expressed and rated their strength on a scale of low to high. The analysis of the results examined both the extent to which expressive intent in a musical performance was perceived as a function of the melody and the instrument playing it, and what timbre's role is in this communication process. Results showed that although there was a communication of affective information, the musical communication process first and foremost depended heavily upon the excerpt, and then additionally, the instrument and/or instrumentalist, and then the intended emotion.

Signifying Emotion: Determining the relative contributions of cues to emotional communication

Aimee Battcock[1], Diana Martinez[2], Michael Schutz[1]
[1]McMaster Institute for Music and the Mind. Hamilton, ON, CANADA
[2]Loyola Law School. Los Angeles, CA, USA

Research on conveyed emotion in music often focuses on experimentally manipulated or newly composed excerpts. Few studies utilize unaltered musical passages composed by internationally acclaimed musicians, in part due to challenges equating variables of interest in complex, natural musical passages. Here we examine the relationship between composers' structural decisions across musical styles (preludes vs. fugues) and listeners' evaluations of conveyed emotion. We focused on 3 musical cues known to play a key role in communicated emotion: mode, pitch height and timing (in the form of attack rate). Non-musician participants listened to 8-measure excerpts of each of the 24 preludes and 24 fugues in Bach's *Well Tempered-Clavier*. After each excerpt, participants rated perceived emotion along a scale of valence and intensity.

Regression analyses revealed timing, pitch height and mode were predictive for mean ratings of valence among preludes (R^2=0.82) and fugues (R^2=0.86). Timing accounted for a larger amount of variance across both musical styles (62%), while pitch height and modality accounted for low to moderate variance in both preludes and fugues. Mean intensity ratings were also significantly predicted by timing within preludes and fugues, however pitch height and modality were not significant. Therefore, it appears timing was the strongest predictor of participants' emotional responses. Cue strength however, varied with respect to style; attack rate accounted for 66% of variance in emotional responses to fugues, and comparatively 76% in responses to preludes. In addition both pitch height and modality cues explained more variance for preludes than fugues. Overall these three cues accounted for more variability within ratings for preludes (R^2=0.80) performance than fugues (R^2=0.68). We will further discuss implications of these differences and the importance of timing as a cue on understanding the communication of emotion in composed music.

Neurophysiological Investigation of Context Modulation based on Musical Stimulus

Siddharth Mehrotra, Anuj Shukla, Dipanjan Roy

Cognitive Science Lab, International Institute of Information Technology, Hyderabad, India.

siddharth.mehrotra@research.iiit.ac.in, anuj.shukla@research.iiit.ac.in, dipanjan.roy@iiit.ac.in

ABSTRACT

There are numerous studies which suggest that perhaps music is truly the language of emotions. Music seems to have an almost willful, evasive quality, defying simple explanation, and indeed requires deeper neurophysiological investigations to gain a better understanding. The current study makes an attempt in that direction to explore the effect of context on music perception. To investigate the same, we measured Galvanic Skin Responses (GSR) and self-reported emotion on 18 participants while listening to different Ragas (musical stimulus) composed of different Rasa's (emotional expression) in the different context (Neutral, Pleasant, and Unpleasant). The IAPS pictures were used to induce the emotional context in participants. Our results from this study suggest that the context can modulate emotional response in music perception but only for a shorter time scale. Interestingly, here we demonstrate by combining GSR and self-reports that this effect gradually vanishes over time and shows emotional adaptation irrespective of context. The overall findings suggest that specific context effects of music perception are transitory in nature and gets saturated on a longer time scale.

I. INTRODUCTION

According to (Levitin, 2006), music is an organized sound, but the organization has to involve some element of the unexpected or it is emotionally flat and robotic. The appreciation we have for music is closely related to our ability to learn the underlying structure of the music we like- the equivalent to grammar in spoken or sign languages- and to be able to make predictions about what will come next. The thrills, chills and tears we experience from music are the results of having our expectations artfully manipulated by a skilled composer and the musicians who interpret the music.

There are three philosophical problems on music and emotion. The first issue concerns the fact that music is frequently perceived as a manifestation of various emotions – despite the fact that music is not a sentient being. The second issue involves the listener's emotional response to music in such instances where the listener's response mirror's the music's, despite a lack of beliefs that usually underline such a response. The third issue concerns why do a listener enjoy and revisit pieces of music that, on their own account, incline them to feel sad or happy. On examining the major philosophical theories of music such as expression theory, arousal theory, music as a virtual persona and contour theory which involves context which is attached to the music (Sharpe, 2004 & Ridley, 2004).

Music is a widely available form of media with the ability to influence attitudes and manipulate emotions (Juslin, 2008 & Wheeler, 2001) and listeners are drawn to music that reflects or improves their emotional state (Saarikallio, 2011; Thoma et al., 2012 & Papinczak et al., 2015). In Indian Classical Music, the melodic basis of compositions and improvisations is based on distinct melodic modes, called 'Ragas.' The etymological meaning of the word 'Raga' is derived from a Sanskrit word 'Ranjayati iti raga', meaning 'that which colours the mind.' Ragas are closely associated with particular emotional themes, termed as 'rasas' (emotional essence). Expression of the raga-rasa aspect is considered the primary goal in Indian Classical Music and this expression is intended to vary dynamically during a piece's performance.

The impact of music on the human body is a significant trend in music research. Different kinds of music have direct and indirect effects on physiological functions e.g. heart rate in normal and pathological conditions. Among various physiological measurements, the galvanic skin response is a noninvasive, useful, simple and reproducible method of capturing the autonomic nerve response (Vanderark, 1992). We used self-reports and GSR (Galvanic Skin Response) responses by participants as a measure to study the emotional adaptation over time while listening to ragas of Indian Classical Music. Our specific hypothesis is that adaptation will take place over time irrespective of emotional context. To tap into the adaptation effect, we will be mainly focusing on the GSR potentials and its saturation.

II. MATERIAL AND METHODS

A. Participants

Eighteen university students, 10 men ($M = 23.3$, $SD = 3.66$) and 8 women ($M= 22.9$, $SD = 5.43$), took part in the experiment. All participants were musically untrained and reported normal hearing. Informed consent was obtained from all participants. The Local Human Ethics Committee of the International Institute of Information Technology, India, approved the study.

B. Apparatus

Electrodermal Activity was recorded using BIOPAC MP150 System with PPGED-R configuration. The sampling rate was 100 samples/second. EDA measurements were recorded continuously during the entire course of the experiment. The experiment was designed using PsychoPy Software (Peirce, 2007).

C. Stimulus

Visual Stimulus: Visual stimuli were presented for 1.5 s in an event-related design. Three categories of stimuli were displayed from the International Affective Picture System (Lang et al., 2008). The images (pixel dimensions: 400X 300) were presented in 32-bit color (resolution: 1024X768 pixels).

i. Pleasant Group: 15 pleasant images were selected from IAPS images database. The valance was equal to or greater than 5 (mean pleasure rating = 7.05, SD = 0.63, range = 5.008.34; mean arousal rating = 4.87, SD = 0.98, range = 2.907.35).

ii. Unpleasant Group: 15 unpleasant images were selected from IAPS images database. The valance was less than the neutral midpoint of 5 (mean pleasure rating = 3.05, SD = 0.84, range = 1.454.59; mean arousal rating = 5.56, SD = 0.92, range = 2.637.35).

iii. Neutral Group: This group served as a Control Group. Subjects were shown 15 images accounting arousal and valence of range 3.5-4.5 based on IAPS normative ratings. All the IAPS pictures were classified as per recommendation of (Mikels et al., 2015).

Musical Stimulus: A set of 12 different ragas were taken from SwarGanga Music Foundation. The duration of each raga was of 3 minutes and were classified as Happy, Sad, Angry and Calm based on their rasa (see Table 1).

Table 1. This table lists the *ragas* used in the study and the *rasa* used by the artist to play the *raga*

Rasa	***Raga***
Hasya (Happy)	Bhupali (Bansuri-flute) Kamaj (Sitar-stringed) Bhupali (Surbahar)
Karuna (Sadness)	Jogya (Dilruba-stringed) Bhairavi (Bansuri-flute) Bhopali-todi (Bansuri-flute)
Raudra (Anger)	Hindol (Bansuri-flute) Adana (Sitar-stringed) Sohini (Bansuri-flute)
Shanta (Calm)	Yaman Kalyan (Bansuri-flute) Yaman (Surbahar-stringed) Bhilaskhani (Dilruba-stringed)

Therefore, in total there were 3 ragas of the same rasa. The above classification was made in accordance with previous literature by (Balkwill & Thompson, 1999; Mathur et al., 2015). The musical structure and the sequences chosen are essentially fragments of ragas. Ragas were played randomly to each participant and participant were asked to report their Mood, Emotion and Tension level on a 5 point likert scale. Mood was rated on a 1 to 5 scale where 1 being extremely annoyed, 2 being annoyed, 3 being neutral, 4 being appeased and 5 being extremely appeased. Emotion was also rated on a 1 to 5 scale where 1 being extremely sad, 2 being sad, 3 being neutral, 4 being happy and 5 being extremely happy. Lastly, for tension Level the scale was 1 being extremely tensed, 2 being tensed, 3 being neutral, 4 being relaxed and 5 being extremely relaxed.

III. PROCEDURE

Participants were tested individually in a dimly lit soundproof room, approximately 57 cm from a 17.8' X 14.2' inches computer monitor. The loudness was set at a comfortable level by the participants at the beginning of the experiment. When the participants arrived at the laboratory, they were given a brief idea about the protocol related to the GSR recording so that the they understand the procedure and artifact can be reduced.

Participants were explicitly instructed to avoid hand movement as the electrodes were placed on index and middle finger. They were also instructed on screen which steps to follow, e.g. pressing spacebar after a visual stimulus, directions of giving ratings etc. The electrodes were attached to the participants 10-15 minutes before the experiment to record their baseline skin conductance. After receiving the instructions and having their electrodes attached, participants rated their initial mood. Participants were asked to report their emotional quotient on 3 categories namely mood, emotion, tension. Most of the participants rated their emotional quotient as calm before going on the main experiment. All the participants were randomly assigned into 3 groups namely pleasant, unpleasant and neutral and presented different emotional pictures depending on the groups e.g. in the pleasant group, participants were shown 15 pleasant picture and again asked them to rate their mood to make sure that participants should be in a particular emotional context and then they were allowed to take part in the main experiment.

After rating their emotional quotient within their respective groups, subjects were blindfolded and were given BOSE 23XSS noise cancellation headphones to listen to the musical stimuli. Ragas were presented randomly and participants provided their responses after listening to each raga for 3 minutes. We collected self-reports in between the two musical stimuli in order to minimize interference with ongoing measurement of electrodermal activity. Consequently, after the experiment, participants were debriefed about scope of experiment.

IV. RESULTS

The analysis was performed in two steps: 1) Pre-processing of GSR Data and 2) Performing Statistical Analysis. GSR data was firstly down-sampled to 25Hz to perform adaptive smoothing. Adaptive smoothing on data was carried out to ensure that artefact removal does not cause distortions in the data. GSR data was then divided into chunks of time-blocks for each raga that was played with the use of marker embedded by PsychoPy software.

The GSR data were analysed using two-way ANOVA with the effect of context (group) on musical stimulus as between-subjects factor and effect of rasa of the raga as within-subject factors. The analysis was performed in MATLAB 2014b and SPSS version 16.0.

Results show a significant effect of group (Pleasant, Unpleasant and Neutral) in which participants were placed ($F(2, 60) = 3.874$, $p = 0.026$) for first 60 seconds which is also reported in our previous findings (Mehrotra et al., *in submission*). An interaction effect was also observed among group and ragas ($F(6, 60) = 2.585$, $p = 0.027$) for first 60 seconds. However, while considering 60 seconds to 120

Table of Contents for this manuscript

seconds, we did not find any significance of the group in which participants were placed (F (2, 60) = 1.686, p = 0.194).

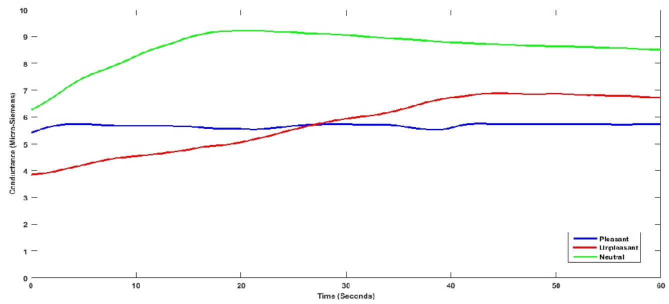


Figure 1. GSR Potential: Angry *Raga* for *60-120 Secs* time interval

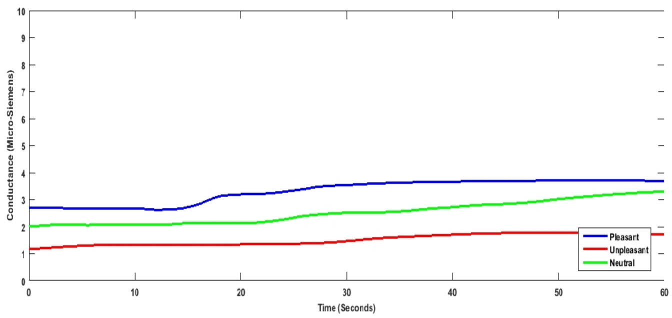


Figure 2. GSR Potential: Calm *Raga* for *60-120 Secs* time interval

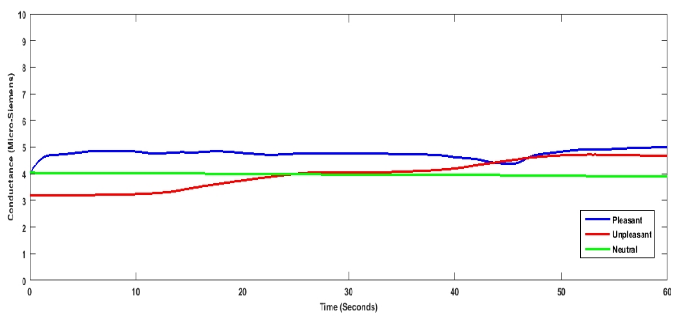


Figure 3. GSR Potential: Happy *Raga* for *60-120 Secs* time interval

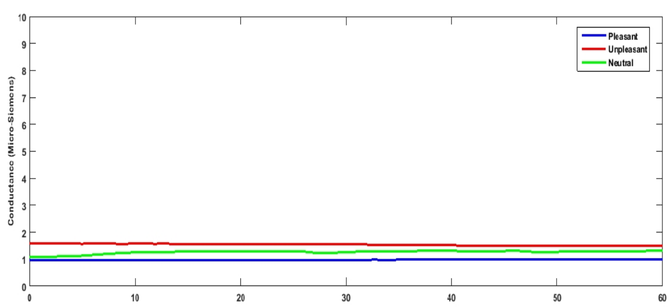


Figure 4. GSR Potential: Sad *Raga* for *60-120 Secs* time interval

A similar trend was observed for 120-180 seconds (F (2, 60) = 1.356, p = 0.265). In figure 2 and 3, we can see that the context has no dependence on the emotional group. For 60-120 seconds, in figure 1, Neutral group has maximum average GSR potential followed by Unpleasant and Pleasant group for angry raga but at the end of 120 secs, it appears that difference in the GSR potential between Neutral and Unpleasant has reduced and hence points towards the adaptation takes place across the group. Further, this observation compliments the statistical result. We observe a similar trend for the other ragas.

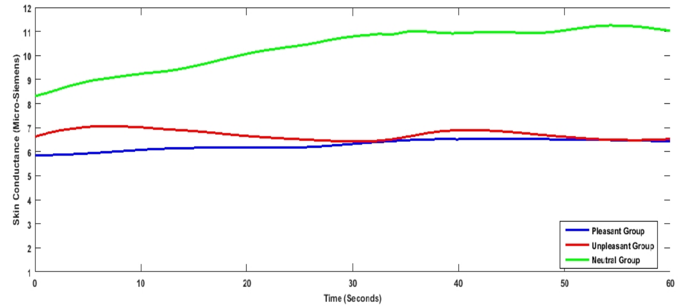


Figure 5. GSR Potential: Angry *Raga* for *120-180 Secs* time interval

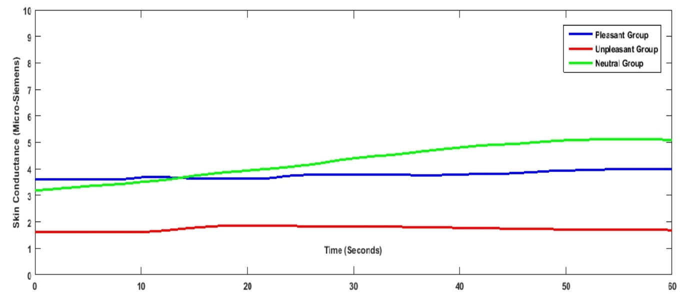


Figure 6. GSR Potential: Calm *Raga* for *120-180 Secs* time interval

For 120-180 seconds, the pleasant and unpleasant group have relatively same average GSR potential followed by the neutral group for angry raga in figure 5. In figure 6, from the continuation of figure 2, Neutral group has achieved highest GSR potential followed by Pleasant and Unpleasant group for Calm Raga. In figure 7, we can follow the same continuation of a trend from figure 3. The Pleasant group has a maximum GSR potential closely followed by the Unpleasant group and the minimum GSR potential is exhibited by the Neutral group for Happy raga. We interpret this as an adaption effect of the musical stimuli that is reaching a plateau over time. Adaption effect is found in both 60-120 secs and 120-180 secs for Calm, Sad, Angry and Happy raga.

In figure 8, the Unpleasant group has highest GSR potential followed by Neutral group and Pleasant group for sad raga. Sad raga shows the strongest effect of adaptation for all groups viz. the Pleasant, Unpleasant and Neutral group for the musical stimulus.

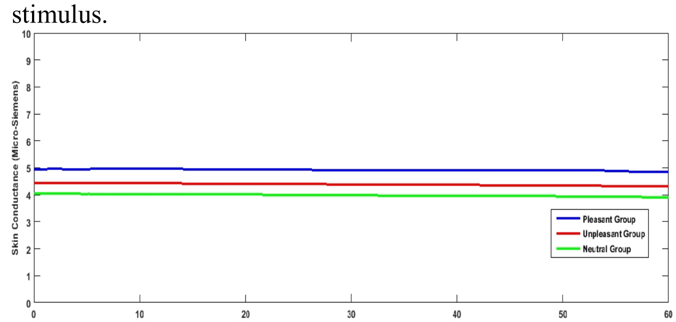


Figure 7. GSR Potential: Happy *Raga* for *120-180 Secs* time interval

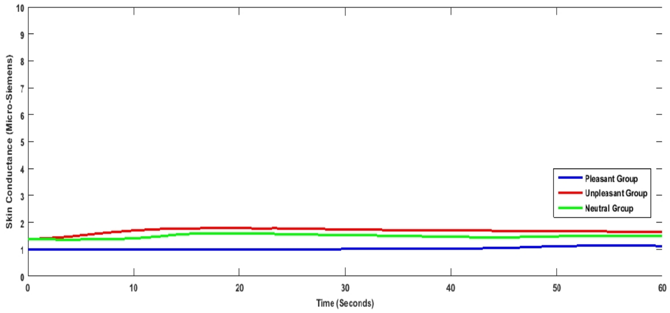


Figure 8. GSR Potential: Sad *Raga* for *120-180 Secs* time interval

V. CONCLUSION

In our previous study, we have established that as the context varies, the emotion in participants also changes despite expression of the music i.e. rasa. The context plays an important role in the perception of Ragas. However, this change is temporal in nature. In the current study, our findings suggest that this effect vanishes over time and adaptation take place. We found an adaptation effect towards the musical stimulus for all participants irrespective of the group in which they are placed. The current finding provides new insight into the music perception studies and indicates that our emotional responses may not be consistent over time for the given music and might have persistence of adaptation effect. Further, the current finding needs investigation and further validation across cultures to gain more insight.

ACKNOWLEDGMENT

We are grateful to all our participants for participating in the study. We are thankful to Prof. S.Bapi Raju for their helpful suggestions. This work was funded in part by the Kohli Center on Intelligent Systems (KCIS), as well as the International Institute of Information Technology, India.

REFERENCES

Balkwill, L., & Thompson, W. F. (1999). A Cross-Cultural Investigation of the Perception of Emotion in Music: Psychophysical and Cultural Cues. *Music Perception: An Interdisciplinary Journal*, 17(1), 43-64

Carifio, J., & Perla, R. J. (2007). Ten Common Misunderstandings, Misconceptions, Persistent Myths and Urban Legends about Likert Scales and Likert Response Formats and their Antidotes. *Journal of Social Sciences,* 106-116. doi:10.3844/jssp.2007.106.116

Dade, P. (2010). A Dictionary of Psychology (3rd ed.)201010 Andrew M. Colman. *A Dictionary of Psychology* (3rd ed.) Oxford: Oxford University Press 2009. xi 882 pp., ISBN: 978 0 19 953406 7

Huron, D. (2003). Is Music an Evolutionary Adaptation*? The Cognitive Neuroscience of Music*, 57-75.

Lang, P.J., Bradley, M.M., & Cuthbert, B.N. (2008*). International affective picture system (IAPS)*: Affective ratings of pictures and instruction manual. Technical Report A-8. University of Florida, Gainesville, FL.

Levitin, D. (2007). *This is Your Brain on Music*. By Daniel Levitin. London: Atlantic Books, 2007. 320 pp. ISBN 978-1-84354-715-0 (hb).

Mathur, A., Vijayakumar, S. H., Chakrabarti, B., & Singh, N. C. (2015). Emotional responses to Hindustani raga music: The role of musical structure. *Frontiers in Psychology Front. Psychol., 6*. doi:10.3389/fpsyg.2015.00513

Mehrotra, S., Shukla, A. & Roy D., Effect of context on perception of *Ragas* (2016). *Psychology of Music* (submitted), SAGE publications

Mikels, J. A., Fredrickson, B. L., Larkin, G. R., Lindberg, C. M., Maglio, S. J., & Reuter-Lorenz, P. A. (2005). Emotional category data on images from the international affective picture system. *Behavior Research Methods, 37*(4), 626-630. doi:10.3758/bf03192732

Papinczak, Z. E., Dingle, G. A., Stoyanov, S., Hides, L., and Zelenko, O. (2015). Young people's use of music for well-being. *J. Youth Stud.*doi:10.1080/13676261.2015.1020935

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1-2), 8-13. doi:10.1016/j.jneumeth.2006.11.01

Peretz, I., & Zatorre, R. J. (2003). *The cognitive neuroscience of music*. Oxford: Oxford University Press.

Saarikallio, S. (2011). Music as emotional self-regulation throughout adulthood. *Psychol.Music* 39,307–327.

Sharpe, R. A. (2004). *Philosophy of music*. Montréal: McGill-Queen's University Press.

Swar Ganga Music Foundation, http://www.swarganga.org

Thoma, M. V., Ryf, S., Mohiyeddini, C., Ehlert, U., and Nater, U. M. (2012). Emotion regulation through listening to music in everyday situation. *Cog.Emot.,*26,50

Wheeler, B. L. (n.d.). *Music therapy handbook.*Wheeler, B. L. (n.d.). *Music therapy handbook*.

The ‘*rasa*’ in the ‘*raga*’? Brain networks of emotion responses to North Indian Classical *ragas*

Avantika Mathur[1] and Nandini Chatterjee Singh[1]
[1]Language, Literacy and Music Laboratory, National Brain Research Centre, NH-8, Manesar – 122 051, India, email: {avantika@nbrc.ac.in, nandini@nbrc.ac.in}

The central notion in North Indian Classical Music is that of a *raga*. The word ‘*raga*’ originates in Sanskrit and is defined as 'the act of colouring or dyeing' (the mind and mood/emotions in this context) and therefore refers metaphorically to 'any feeling of love, affection, sympathy, desire, interest, motivation, joy, or delight’. It is believed that an artist uses particular note combinations, to create a mood (*rasa*) that is unique to the *raga*. We conducted a behavioural study in which 122 participants from India rated their experienced emotion for 12 *ragas* on Likert scale of 0-4 (with 0 being ‘not at all felt’ to 4 being ‘felt the most’) for each of the following emotions - happy, romantic, devotional, calm, angry, longing, tensed, and sad. The results of the behavioural study conducted online revealed that *ragas* were indeed capable of eliciting distinct emotions. Eight ragas that were rated highest on ‘happy’ and ‘tensed’ emotion, four in each category were chosen for a subsequent functional neuroimaging experiment to investigate the neural circuits underlying these emotional responses. A separate group of 29 participants rated *raga* sequences selected from the behavioural study for ‘happy’ and ‘tensed’ emotion. The *raga* condition contrasted with rest revealed a network of areas implicated in emotion processing namely, bilateral mOFC, ACC, caudate, nucleus accumbens, insula, precuneus, auditory association areas and right lOFC, amygdala, hippocampus - para hippocampus gyrus complex and hypothalamus. Further flexible factorial analysis revealed a main effect of emotion (happy/tensed) in the bilateral mOFC. Our results provide evidence for the first time that *ragas* elicit basic emotion circuitry. They also reveal that valence information may be coded in the mOFC. Together, we demonstrate the ability of North Indian Classical *ragas* as tone sequences capable of eliciting distinct emotional responses.

Multimodal affective interactions: Comparing auditory and visual components in dramatic scenes

Moe Touizrar[1], Sarah Gates[1,2], Kit Soden[1], Bennett K. Smith[1], Stephen McAdams[1]
[1]Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT), Schulich School of Music, McGill University, Montreal, QC, CANADA
[2]now at Bienen School of Music, Northwestern University, Evanston, Illinois, USA

Links between music and emotion in dramatic art works (film, opera or other theatrical forms) have long been recognized as integral to a work's success. A key question concerns how the auditory and visual components in dramatic works correspond and interact in the elicitation of affect. We aim to identify and isolate the components of dramatic music that are able to clearly represent basic emotional categories using empirical methods to isolate and measure auditory and visual interactions. By separating visual and audio components, we are able to control for, coordinate and compare their individual contributions. Using stimuli from opera and film noir, we collected a rich set of data (N=120) that will make it possible for us to segment and annotate musical scores with collated affect descriptors for subsequent score analysis. Custom software was developed that enables participants successively to record real-time emotional intensity, to create event segmentations and to apply overlapping, multi-level affect labels to stimuli in audio, visual and audiovisual conditions. Our findings suggest that intensity profiles across conditions are similar, but that the auditory component is rated stronger than the video or audiovisual components. Descriptor data show congruency in responses based on six basic emotion categories, and suggest that the auditory component elicits a wider array of affective responses, whereas the video and audiovisual conditions elicit more consolidated responses. These data will enable a new type of score analysis based entirely on emotional intensity and emotion category, with applications in music perception, music theory, composition, and musicology.

Differences in Verbal Description of Music Listening Experiences between College Students with Total Blindness and Typical Vision

Hye Young Park,*[1] Hyun Ju Chong,#[2]

[1]Music Therapy, Kosin University, Busan, South Korea

[2] Department of Music Therapy, Ewha Womans University, Seoul, South Korea

[1]legendphy@kosin.ac.kr, [2]hju@ewha.ac.kr

ABSTRACT

The aim of this study is to examine the types of verbal expressions used by college students with and without visual impairments. Participants of the in-depth interviews were students at universities in Seoul and Chungcheong Province. Ten participants were visually impaired and the other ten were normal sighted. Content analysis of the in-depth interviews revealed the following implications. First, visually oriented verbalism is observed among visually impaired people even when they can experience the same level of musical experience with normal sighted people. Highly observed is verbalism by contextual information which is naturally obtained in daily life, rather than verbalism gained from the media nor from didactic learning. Verbalism is a linguistic characteristic shown by the people with visually impaired using words in an abstract way. Second, K-LIWC was employed to examine the types of words used for description between the two comparing groups, and results showed high ratio of use of pronouns in the visually impaired was observed. Third, In terms of the content, words for negative emotions were used more frequently. Fourth, compared to normal sighted people, visually impaired people more frequently use affective description and description via senses other than vision which implies that the people with visually impairment use non-vision senses substitute or compensate visual impairment. Last, in regards to episodic memory and situational depiction, normal sighted people frequently narrate episodic memories, while people with visually impairment more frequently focus on situational depiction.

I. INTRODUCTION

Visual imagery is one of major mechanism to induce music emotion. It has been found that the people with visually impaired (VI) make less analogy or metaphorical expression than those with normal vision (NV) due to the listener's ability of processing music and organizing them into visual images. Visual experience limitation affects abstract information processing in music listening experience.

The factors which influence on VI's listening of music include conceptualization of music and visual imagery process (Darrow & Novak, 2007). While there are studies which suggest that VI's tactile sense coincides with normal sighted people's visual sense (Walker, 1985), the studies on different amount of imagery utilized by visually impaired and normal sighted people using the similar measuring tools suggest that limited visual experience has negative impacts on the formation of musical conceptualization for VI (Madsen & Darrow, 1989).

Also, the study which revealed limitation on their process of abstract data, reported that when asked to freely comment about the music after listening, the VI tend to present more data-driven description including used instruments and musical factors while presenting less of analogical or metaphorical descriptions comparing to the normal sighted (Flowers & Wang, 2002). Despite their visual impairment they can still articulate the music listening experiences in their own words.

Meanwhile, despite the fact that language is less affected by visual impairment since it is learned through listening (Anderson & Olson, 1981; Civelli, 1983), it is hard to find studies analyzing the VI's verbal description of emotional reaction upon listening to music. Since the main route of communication with others of the visually impaired is listening, it is necessary to pay attention to verbal description of music after listening in order to look into music listening experience of the visually impaired and the normal sighted.

Research on the influence of visual impairments on verbal description in everyday life has been continued by the scholars majoring in visual impairment education. Regarding linguistic characteristic caused by the loss of vision, Cutsforth(1951) used the word verbalism for the first time, and this word means the linguistic characteristic of the visually impaired in which he/she uses the word which he/she heard and did not fully understand in an abstract way in everyday life. Especially, visually oriented verbalism refers to congenital VI's behavior of using the words related to visual information such as color, shade, chroma in everyday life without having any direct experience of them (Vinter, Fernandes, Orlandi, & Morgan, 2013). This is closely related to the age as it is found that there is tendency for the verbal similarity between the visually impaired and the normal sighted to increase as they get older. (Rosel et al., 2005).

This study purported to investigate the differences of music listening experiences between adults with normal sight and visual impairment focusing on their characteristics of verbal descriptions. The research questions were:

First, what are the differences between descriptive factors of two different groups of participants after the music listening experience?

Second, what are the differences between perceptual factors of two different groups of participants after the music listening experience?

Third, what are the differences between referential factors of two different groups of participants after the music listening experience?

II. METHOD

Participants (10 VI, 10 NV) were students at five universities in Korea. Self-volunteered participants listened to the selected music for approximately 3 minutes, and freely described their emotions, physical reaction, related images, episodic memory, specific emotion related to musical factor from listening to music. For the in-depth interview, the authors played the selected emotion-inducing-music to the participant, and the participant freely expressed and described the emotions he/she had felt using adjectives, and the authors recorded this process. The authors collected all the raw recorded data, systematically organized them from interviews and observations, and documented them. For the additional K-LIWC to be held, written interview contents were saved as computer text files. This stage was progressed as the 1 on 1 interview format, and the running time was within 50 minutes.

The procedures which took place for the analysis of verbal description about music listening experience were as follows: 1. The authors listened to the recorded data multiple times, analyzed them by lines, and went through the classification process by grouping them by conceptual bases or occurring topics. After, classified lines and phrases were repeatedly classified, and mutually exclusive 3 high ranked sources were decided based on the results. Also, based on the characteristic of each high ranked source, if the category of the source seemed broad, low ranked sources were classified based on the smaller topics. 2. Coding process took place, categorizing and organizing previously drawn category's lines and phrases. Since author's subjectivity could be involved in the process, there were two processes added in which the doctoral student each encoded and compared the results to verify reliability and validity of the content analysis.

Content analysis emphasizes repeatability in the process of drawing inference from the data (Krippendorff, 2004), thus, obtaining reliability of the study, and conformity degree between the coders in the encoding process were reviewed as part of the whole process. The result of the first encoding verification between the authors and doctoral student revealed that 108 of 126 showed similar category, showing 86.1% of conformity degree. For the sections which showed different category, the author and doctoral student went through discussion and agreement steps in order to carry out second encoding process, and finally reached 91.4% of conformity degree. Verbal description regarding the categorized emotional reaction of music has been converted into quantitative data and has been presented by frequency. Also, in order to supplement any contents that may be overlooked by processing it by quantitative data, qualitative content analysis was also carried out, and quotations for each category were presented as the example.

In order to find out about descriptive factors in the interview contents, analysis took place in the psychological and syntactic formats utilizing K-LIWC. The analyzing process included proof reading of the text file of the interview of research participants for any grammatical or spelling errors, and analyzing of the text file with online K-LIWC analysis program. After running the morpheme analysis by tagging system, a post-processing-system process was held with concept & vocabulary establishment system. In this process, the authors performed correction operation on errors of morpheme and ambiguous significance. Through the above operation processes, result files have been drawn on psychological and linguistic factors. During this research, 6 causes following syntactic format in linguistic factor have been reviewed, and optimistic & pessimistic emotional process, recognition process, and psychological dimension about perceptual process have been analyzed as the 8 causes. In order to secure credibility of this study, methods such as triangulation, peer debriefing, member checking were employed (Stemler, 2001).

III. RESULTS

Content analysis of the in-depth interviews revealed the following results. The result of analysis of linguistic description regarding music listening experience was divided into three types based on the research problem. First, in order to look into descriptive factors regarding musical emotion reaction between two groups of participants, visually oriented verbalism which examines characteristics of congenitally blind college students and form of description were reviewed. Especially in the descriptive factors, validity of the result was secured as analysis through K-LIWC was included along with content analysis. Second, in order to analyze perceptual factor of musical emotion reaction between two groups of participants, lower ranked category of visual, sensual, emotional was used. Third, in the research regarding referential factor item of musical emotional reaction between two groups of participants, only musically induced anecdote remembrance and situational description were analyzed excluding any contents with direct reference of musical terms.

A. Visually Oriented Verbalism

In this study, the behavior of totally blind college students, in which they would describe color, brightness, size, shape as if they had personally experienced such types, was categorized as visually-oriented-verbalism (Vinter et al., 2013). Visually oriented verbalism was classified into three types including media, didactic learning, and contextual use based on the source of visual information. Verbalism through media refers to imitated description of visual information obtained from media including novels, animations, movies or television shows, and verbalism through didactic learning refers to imitated description of visual information learned through education or teaching environment. In addition, contextual verbalism refers to inferred description of naturally obtained visual information from flow, atmosphere, and scene of other's linguistic expression (see Table 1).

<Table 1> Analysis of visually oriented verbalism regarding music listening experience (*n*=10)

Category	Subcategory	Concepts	*n* (%)
Visually oriented verbalism	Verbalism through didactic learning	Teacher explains the shape and color of the instrument Parents explains the clothing shapes and color Friend describes the scenery of the sunset Learns about the color of the food through everyday training material Helper who went to the concert together explains the view performers playing instruments	7(20.0)
	Verbalism through Media	Read about the portrayal of seasonal sense of color in the novel Learns about the image of the nature through descriptive caption of the movie	8(22.8)

		Understands about atmosphere of specific place through TV drama Learns about facial expression and attitude through lyrics of the song	
	Contextual verbalism	Understands about symbols of color through conversation with friends Infers about the size of the object through hearing about portrayal of specific scene Learns about the change of light through situational description Understands about characteristic of the object by combining survival sense and visual information	20(57.2)
Total			35(100)

B. Form of Description

Form of description was analyzed by breaking down interview recording in linguistic dimension and psychological dimension using Korean-Linguistic Inquiry Word Counting (K-LIWC). By looking into syntactic aspect including word, morpheme, word vs. sentence, it was intended to find out additional information about totally blind college students use of words. Linguistic dimension is the method to understand the characteristic of use of words of totally blind college student through syntactic approach. It includes major linguistic dimension's 6 low ranked factors including noun, pronoun, verb, adjective, adverb, and exclamation (see Table 2).

<Table 2> Linguistic characteristic of descriptive factor about music listening experience (*N*=20)

Types	VI (*n*=10)	NV (*n*=10)	*z*	*p*
Noun	14.6	15.1	-1.369	0.17
Pronoun	1.8	0.7	-2.389	**0.04***
Verb	10.8	13.3	-1.579	0.11
Adjective	5.5	5.7	-0.684	0.49
Adverb	2.9	2.6	-0.526	0.59
Exclamation	0.8	0.3	-0.919	0.35

$^{*}p < .05$

C. Perceptual Factor

Since this study is the comparative analysis of linguistic description characteristic about emotional reaction triggered by music between totally blind college student and normal sighted college student, visual adjective has been categorized as a separate category, while other senses including hearing, smell, and touch have been categorized as one category as a whole. The words included in the emotional category refer to all the words that could describe the reaction from external stimulation (see Table 3).

<Table 3> Analysis of perceptual factor content of music listening experience

Category	Subcategory	VI (%)	NV (%)	Total (%)
Perceptual Factor	Visual	52 (16.0)	78 (25.2)	130 (20.4)
	Sensual	99 (30.6)	71 (23.0)	170 (26.8)
	Emotional	130 (40.3)	95 (30.6)	225 (35.6)
	Behavioral	42 (13.1)	66 (21.2)	108 (17.2)
Total		323 (100)	310 (100)	633 (100)

D. Referential Factor

The referential factor analysis unit in the music listening experience is sentence, and it needs to have subject and predicate. Also, even if the subject is excluded in the special case of interview, if the meaning is clear, then it would consider it as the analysis subject. Episodic memory triggered by music is the description about research participant's experienced incident, people, and place. In addition, situational depiction is the description about story telling or imaginary image triggered by music (see Table 4).

<Table 4> Analysis of referential factor about music listening experience (*N*=20)

Category	Subcategory	VI (%)	NV (%)	Total (%)
Referential Factor	Episodic Memory	20(35.0)	27(49.1)	47(41.9)
	Situational Depiction	37(64.8)	28(50.9)	65(58.1)
Total		57 (100)	55 (100)	112 (100)

IV. CONCLUSION

The conclusion derived from the analysis of verbal description about music listening is as follows.

First, VI's showed high level of visual directivity verbalism, and its visual information was mostly acquired from contextual verbalism, followed by verbalism acquired from media and didactic learning. This implies how use of verbalism itself has contextual characteristics, and shows how information obtained from everyday life or conversation with others have bigger impacts than those obtained from books, movies, television or through education. In other words, it proves that VI tend to have higher trust on visual information obtained contextually compared to that from meda or education, and tend to utilize them as one of their ordinary verbal or linguistic habits (Jaworska-Biskup, 2011). Generally, amount of contextually obtained linguistic habit through languages used by people is much more than that obtained from education or media, and there is higher trust for contextually obtained visual information (Kemp, 1981). In other words, as it is used continuously in the casual conversation with others, it is silently approved, and accepted contextual verbalism is used frequently as way of communication applying the characteristic of language.

Secondly, through the Korean-Linguistic Inquiry Word Counting (K-LIWC)'s syntactic factor analysis in order to find the differences in the description forms, it was found VI group tend to frequently use pronouns (Perez-Pereira, 1999). This could be seen that the concentration level of VI was higher than that of NV, and they believed that their use of pronoun would imply the actual noun it is referring to the interviewer. Another assumption is that it could be just another method of communication for them to describe unclear contents (Andersen, Dunlea, & Kekelis, 1984). This could be supported by the characteristics shown in the overall talking pattern. If you look at the interview data, it shows that totally blind college students tend to use longer sentences compared to normal sighted college students, and showed characteristics of frequent use of overlapping words. The reason for totally blind college student to show higher frequency use of pronoun is because when trying to come up with linguistic description of emotions derived from music, normal sighted college students are able to express detailed visual or imagery description, while totally blind college students need to use pronouns more often when subject is unclear because there are many imaginary and situational descriptions. However, in the psychological factor analysis of K-LIWC, it was found that there was a higher negative emotion for totally blind college students while there were higher figures in cognitive process for normal sighted college students. On the other hand, there was not much difference between two groups for the word usage frequency in the perceptual process, and it seems likely that totally blind college students used the perceptual expression as compensation for the subjects he may not be able to recognize visually (Tobin, 1992).

Third, in the perceptual factor category, the description employed affective descriptions for both groups. While NV group used more visual expressions, VI group used more emotional and sensual expressions. This could be attributed to the need to substitute the deficient visual information due to vision-loss with other senses (Eitan, Ornoy, & Granot, 2012). In other words, the VI would obtain visual images that they cannot experience through vision through their other senses, and this acquisition of information through replaced compensation senses shows how it is effectively helping VI to be integrated in society (Park, Chong, & Kim, 2015). The strategy to obtain visual information utilizing other senses deserves much attention as it is a good example of replaced compensation sense different from traditional visually oriented verbalism using contextual characteristic of language.

Lastly, in the referential factor, NV and VI group contrasted significantly in the episodic memory and situational depiction, and it seems that differences in the amount of direct/indirect experience played a significant role in this result. In the episodic memory category, which is based on the specific experience, normal sighted college students seemed to have higher results than totally blind college students because of the differences in the amount of visual experience (Tulving, 2001). The reason for the higher results in the situational depiction category is because although they did not have more experience, they tried harder to express their emotional descriptions in detail through story or event formation (Joseph, & Southcott, 2006). This implies that the music which would arouse visual content or episodic memory for normal sighted people, serves as useful, imaginary, episodic resources for VI to reformulate the situational story. Hence, while visual clues of musical emotions drew direct experience description called episodic memory for the normal sighted, the VI transformed these visual images into situational depiction through the imaginary reformation of the event utilizing other sensual contents.

REFERENCES

Anderson, D. W., & Olson, M. R. (1981). Word Meaning among Congenitally Blind Children. *Journal of Visual Impairment & Blindness, 75*(4), 165-68.

Andersen, E. S., Dunlea, A., & Kekelis, L. S. (1984). Blind children's language. *Journal of Child Language, 11*(3), 645-664.

Civelli, E. M. (1983). Verbalism in young blind children. *Journal of Visual Impairment & Blindness. 2,* 61-63.

Cutsforth, T. D. (1951). *The blind in school and society: A psychological study.* New York: American Foundation for the Blind.

Darrow, A. A., & Novak, J. (2007). The effect of vision and hearing loss on listeners' perception of referential meaning in music. *Journal of Music Therapy, 44*(1), 57-73.

Eitan, Z., Ornoy, E., & Granot, R. Y. (2012). Listening in the dark: Congenital and early blindness and cross-domain mappings in music. *Psychomusicology: Music, Mind, and Brain, 22*(1), 33-45.

Flowers, P. J., & Wang, C. H. (2002). Matching verbal description to music excerpt: The use of language by blind and sighted children. *Journal of Research in Music Education, 50*(3), 202-214.

Jaworska-Biskup, K. (2011). The world without sight. A comparative study of concept understanding in Polish congenitally totally blind and sighted children. *Psychology of Language and Communication, 15*(1), 27-48.

Joseph, D. Y., & Southcott, J. (2006). The blind men and the elephant: music education in a changing world. *Proceedings of the Australian Association for Research in Music Education.*

Kemp, A. (1997). Individual differences in musical behavior. In D. J. Hargreaves & A. C. North (Eds.*), The social psychology of music (pp. 25–45)*. Oxford: Oxford University Press.

Krippendorff, K. (2004). Reliability in content analysis. *Human Communication Research, 30*(3), 411-433.

Madsen, C. K., & Darrow, A. A. (1989). The relationship between music aptitude and sound conceptualization of the visually impaired. *Journal of Music Therapy, 26*(2), 71-78.

Park, H.Y., Chong, H. J., & Kim, S. J. (2015). A comparative study on the attitudes and uses of music by adults with visual impairments. *Journal of Visual Impairment & Blindness, 109*(4), 303-316.

Perez-Pereira, M. (1999). Deixis, personal reference, and the use of pronouns by blind children. *Journal of Child Language, 26*(03), 655-680.

Rosel, J., Caballer, A., Jara, P., & Oliver, J. C. (2005). Verbalism in the narrative language of children who are blind and sighted. *Journal of Visual Impairment & Blindness, 99*(7), 1-22.

Stemler, S. (2001). An overview of content analysis. Practical Assessment. *Research & Evaluation, 7*(17), 137-146.

Tobin, M. J. (1992). The language of blind children: Communication, words, and meanings. *Language and Education, 6*(2), 177-182.

Tulving, E. (2001). Episodic memory and common sense: how far apart? *Philosophical Transactions of the Royal Society B:Biological Sciences, 356*(1413), 1505-1515.

Vinter, A., Fernandes, V., Orlandi, O., & Morgan, P. (2013). Verbaldefinitions of familiar objects in blind children reflect their peculiar perceptual experience. *Child: Care, Health & Development, 39*(6), 856-863.

Walker, A. R. (1985). Mental imagery and musical concepts: Someevidence from the congenitally blind. *Bulletin of the Council for Research in Music Education,* 229-237.

Social eMotions — communicating emotions in contemporary dance

Tommi Himberg[1], Klaus Förger[2], Maija Niinisalo[1], Johanna Nuutinen[3], Jarkko Lehmus[4]
[1]Department of Neuroscience and Biomedical Engineering, Aalto University, Espoo, Finland
[2]Department of Computer Science, Aalto University, Espoo, Finland
[3]Free-lancer, Helsinki, Finland
[4]JoJo, Oulu Dance Centre, Oulu, Finland

Every social interaction is also communication of emotions through our voices, postures, and body movements. Emotions are contagious; they are in a constant flux, and have a big, yet often subconscious effect on our social relations. Dance and music have the ability to bring emotional communication to the foreground, and a wide spectrum of emotions, their intensities, and their social consequences can be explored through them.

We combined contemporary dance and movement science to explore the social, dynamic, and embodied facets of emotions. Two dancers studied social emotions and emotional contagion in their artistic research, and then performed a short choreography with different emotional conditions ("loathing", neutral, "loving", pride & shame; pair having same or different emotions; static, changing together, or one contracting the emotion of the other). The performances were recorded using optical motion capture, and kinematic analysis was conducted to investigate which movement features correspond with which emotional states and scenarios.

Principal component analysis (PCA) of the kinematic features indicates that the first component mainly captures the "vigorousness" of the dance, separating "loving" from "loathing", with neutral performances in between. PC2 captures the social dimension, or movement similarity between dancers. PC3 is loaded with acceleration-related features, and teases apart depictions of pride and shame from the others.

Currently, perceptual experiments to investigate the audience's responses are being conducted. Animations of actual performances and synthesized modifications are used to explore the effect of movement features parametrically.

Our combination of artistic, kinematic and perceptual research on emotion communication allows us to discuss the mechanisms of emotional communication in the context of contemporary dance, and the applicability of them in other contexts, including music performance.

Rehearing music: schematic and veridical expectancy ratings in adults

Hayley Trower[1], Adam Ockelford[1], Evangelos Himonides[2], Arielle Bonneville-Roussy[1]
[1]Applied Music Research Centre, University of Roehampton, London, UK
[2]Department of Culture, Communication and Media, Institute of Education, London, UK

The generation of expectations is integral to the perception and cognition of music; however, the common practice of rehearing music tends to be overlooked. A key question pertaining to expectations that arise from rehearing is: if the violation of expectations communicates emotional pleasure, why does this continue to happen even when the listener knows what is going to come next? The answer may be that there exist two forms of expectation; schematic (learned from an individual's lifelong exposure to music, and always hears music as though for the first time) and veridical (concerning specific memory for a piece or phrase). The role of veridical expectations is downplayed in the literature, despite the intuition that music is often simultaneously surprising and unsurprising. The purpose of the present research is to utilise a continuous response methodology to provide empirical support for the theory that two independent forms of expectation are engaged during listening to familiar music. In a within-subjects experiment, 30 adult musicians and non-musicians gave note-by-note expectancy ratings in response to a 4 x repeated melody. As participants were led to believe that each melody was different, an 8 note distractor was heard between each repetition. Participants rated by moving their finger along a touch sensitive controller known as a Continuous Response Measurement Apparatus or CReMA. An identical experiment was repeated 1 week later. Results support the hypothesis that 2 forms of expectation are activated during music listening. With each melody repetition, participants show a systematic increase in expectedness, yet the pattern of expectation remains intact. This finding suggests that veridical expectations are not supressed by schematic expectations; rather there is a dynamic relationship that alters with each new hearing of a piece of music which enables the listener to repeatedly enjoy a piece of music through the co-existence of two memory systems.

Pleasurable music is serious, exciting and bodily experienced, while pleasurable visual objects are light, calming, and aesthetically pleasing

J. Maksimainen, S. Saarikallio

***Abstract*—**This study is part of a research project exploring the emotional-motivational constitution of everyday pleasure in music and visual modalities. Here we aimed to identify the perceived object characteristics and features of emotional responsiveness constituting the experience of pleasure in music in comparison to visual objects. An online survey (N=464) was used to collect ratings on these aspects. The results showed that pleasurable music was perceived as serious, rebellious, mystical, bold, heavy, exciting, energetic, restless, and sensual; in contrast, pleasurable visual objects were characterized as playful**,** funny, conventional, down to earth, discreet, light, calming, calm, tranquil, and non-erotic. Furthermore, emotional reactivity to music was characterized by strong emotional reactions, increased attention, and bodily experience, while emotional reactivity in visual objects was particularly characterized by aesthetic experience. The study provides novel information about individuals' emotional, pleasure-related attachment to distinct modalities in daily life, laying foundations for the theorization of how people perceive information obtained through these modalities.

***Keywords*—**Music, emotional responsiveness, perception, visuality.

I. Introduction

THIS study is part of a research project that aims to map the emotional-motivational constitution of everyday pleasure induced by music and visual arts. This particular sub-study was designed to identify the perceived object characteristics that constitute the experience of pleasure within music and visual modalities, and to explore the nature of emotional responsiveness towards the pleasurable musical and visual objects. Thus, the main research questions were:

1. What are the similarities and differences in perceived characteristics of pleasurable music vs. visual objects?
2. What are the similarities and differences in emotional responsiveness towards pleasurable music and visual objects?
3. How does emotional responsiveness relate to finding an object pleasurable?

J. Maksimainen is with the University of Jyvaskyla, Department of Music / Department of Art and Culture Studies, Jyvaskyla, Finland (phone: 358-50 411 0678; e-mail: johanna.maksimainen@jyu.fi).

S. Saarikallio is with the University of Jyvaskyla, Department of Music, Jyvaskyla, Finland (e-mail: suvi.saarikallio@jyu.fi).

II. Method

A. Participants

Data were collected through an online survey (N = 464), measuring emotions and concepts related to a musical piece or a visual object that respondents considered pleasure-evoking in their daily life. Participants were recruited though Finnish Universities' mailing lists and social media. Participation was voluntary and participants were not given any incentives. The sample consisted of 464 individuals, with ages ranging from 18 to 82 years (M = 39.9 years, SD = 13.8). The sample was predominantly female (78.9%; 19.2% male; 1.9% other), and highly educated, with a majority of participants reporting a university graduate degree as their highest level of education (14.4% High school examination; 8.6% Polytechnic; 10.8% Bachelor's degree; 39.9% Master's degree; 19.6% PhD or doctorate). The overall sample was divided into two groups, which were self-selected on the basis of reporting either a musical (N = 228; 49.1%) or a visual (N = 236; 50.9%) pleasure-inducing stimulus.

B. Measures

All measures were presented through a semi-structured online questionnaire, which was accessed and responded to by participants using their own personal computing device. Participants were instructed to select one object they considered to induce pleasure and hold personal significance in their daily life. The object could be either musical (a specific song) or visual (a specific object or wider visual environment).

Pleasure-related characteristics of the selected item were assessed by using a Semantic Differential Scale (SDS) of 33 adjective pairs (Appendix 1). The selection of the descriptive concepts to be assessed was influenced by research traditions and approaches from psychology and user experience (e.g. Heise, 2010; Himmelfarb, 1993; Jordan, 2002). The SDS is a tool used frequently for measuring social attitudes (Osgood, Suci & Tannenbaum, 1957). Here it was applied to exploring the perceptions that individuals tend to associate to the object. The SDS is a seven-point bipolar rating scale using opposing adjective pairs from which respondents select a point corresponding to their disposition about the object in question (Osgood, Suci & Tannenbaum, 1957).

The second set of questions related to individuals' *emotional responsiveness* (Appendix 2). The first question assessed the participant's general *Emotional reactivity (general strength of emotional reaction*). The following six questions assessed their emotional responsiveness towards the selected musical or visual object. These consisted of the following aspects of emotional responsiveness*: Emotion strength, Pleasure strength, Aesthetic experience, Attention, Ambivalence,* and *Bodily experience.* Ratings were provided on a 7-point Likert scale (see scale dimensions in Appendix 2). For the analyses, and reporting the results, the scrutinized concepts were translated from Finnish to English using back-translation.

C. Analysis

The analysis was predominantly descriptive and exploratory in nature. We first explored the mean values of the bipolar adjective scales, in order to identify object characteristics and/or associations, and the difference in these between the modalities. In terms of statistical analyses, MANOVA and subsequent t-tests were calculated to detect significant differences between the modalities.

Secondly, similarities and differences between the modalities regarding the seven constituents of emotional responsiveness were investigated by comparing their mean values (MANOVA and subsequent t-tests). Finally, the most distinctive features of the emotional responsiveness in each modality were correlated with the perceived pleasantness of the object.

III. Results

A. Object characteristics

The results showed several distinctions between how pleasurable music and pleasurable visual objects are perceived. First, we searched for such characteristics that were most similarly rated in both modalities. Object characteristics were selected as "similarities" if the mean ratings between the object types were practically identical, i.e. filled the criterion of showing only ≤0.1 unit difference. These characteristics are presented in Table 1. Bolding in Table 1 indicates towards which end of the SDS the mean values in each scale were settled. Both modalities were equally associated to be highly *pleasant* and *timeless,* and somewhat *nostalgic, sophisticated*, and *imposing* (rather than unpleasant, temporal, oriented towards novelty, rough, and modest, respectively).

TABLE I
SIMILARLY PERCEIVED OBJECT CHARACTERISTICS

Object characteristics	Mean / Music (N=228)	Mean / Visual (N=236)
Pleasant - Unpleasant	1.5	1.5
Temporal - **Timeless**	6.4	6.3
Nostalgic - Oriented towards Novelty	3.6	3.5
Rough - **Sophisticated**	4.8	4.9
Modest - **Imposing**	4.6	4.6

Next, we focused on identifying the differences between the modalities. MANOVA indicated a significant difference between the modalities, F (33, 430) = 7.80, p ≤ .001. In order to identify the clearest differences and avoid bias from multiple comparisons, we only selected object characteristics that filled the criterion of having ≥0.5 unit mean difference between the modalities and showing a statistically significant difference at ≤ .001 significance level. These object characteristics are presented in Table 2. Again, bolding indicates towards which end of the SDS the mean values were settled.

TABLE 2
DIFFERENTLY PERCEIVED OBJECT CHARACTERISTICS

Object characteristics:	Mean / Music (N=228)	Mean /Visual (N=236)
Ratings at opposite sides of the SD scale:		
Playful - Serious	4.3	3.5
Conventional - Rebellious	4.2	3.6
Mystical - Down to earth	3.7	4.5
Discreet - Bold	4.5	3.9
Ratings on same side of the SD scale:		
Light - Heavy	3.7	3.0
Calming - Exciting	3.2	2.3
Energetic - **Calm**	4.3	5.0
Funny - **Serious**	4.8	4.0
Tranquil - Restless	2.7	2.1
Sensual - **Non-erotic**	4.0	4.7

"Differences" were categorized in two sub-groups. The first section of Table 2 represents dimensions in which the mean values were generally close to the midpoint of the scale, but towards different ends of the scale in different modalities. Among these, the musical objects inducing pleasure were perceived to be more towards *serious, rebellious, mystical,* and *bold*, whereas visual objects inducing pleasure were perceived to be more towards *playful, conventional, down to earth* and *discreet.*

The second section of Table 2 presents dimensions in which the mean values were settled on the same side of the scale in both modalities, associating both object types as light, calming, calm, serious, tranquil, and non-erotic. However, the mean value differences indicated that the perceived characteristics of music, in comparison to visual domain, were inclined more towards *heavy, exciting, energetic, serious, restless,* and *sensual*. In contrast, characteristics related to visual objects were perceived more towards *light, calming, calm, funny, tranquil and non-erotic*.

B. Emotional Responsiveness

MANOVA revealed a significant difference between the object types in the emotional responsiveness factors, F (7, 239) = 20.43, p ″ .001, and subsequent t-tests were conducted to investigate the differences in each of the seven features (Table 3).

The general tendency for emotional responsiveness (reactivity) appeared to be somewhat stronger within those who selected music in comparison to those who selected a visual object, and an even stronger difference was observed between the modalities for the emotional strength experienced towards the selected object, showing that music was experienced significantly more emotionally than the visual objects. *Pleasure strength* was scored high within both groups (M=6.1/music, M=6.0/visual), and did not show significant difference between the modalities. Specific aspects of emotional reactivity also differed significantly between the object types, showing significantly higher ratings for attention and bodily experience in relation to pleasure evoked by music than for pleasure evoked by visual objects. In contrast, aesthetic experience received significantly higher ratings within visual domain than in music. Mean ratings for the experience of emotional ambivalence were somewhat higher in music than in visual objects but the difference between modalities did not reach significance.

TABLE 3
EMOTIONAL RESPONSIVENESS

Variable		Mean	Std.	t-test results
Reactivity	Music	5.3	1.3	t (462) = 2.19; p=.029
	Visual	5.0	1.4	
EmoStrength	Music	5.7	.90	t (448.67) = 3.82; p ≤ .001
	Visual	5.4	1.1	
PleaStrength	Music	6.1	.91	t (462) = 1.38; ns.
	Visual	6.0	1.0	
AesthExp	Music	5.7	1.5	t (435.33) = -3.49; p ≤ .001
	Visual	6.1	1.2	
Attention	Music	6.0	.92	t (421.39) = 9.12; p ≤ .001
	Visual	5.0	1.4	
Ambivalence	Music	3.2	2.0	t (449.72) = 1.93; ns
	Visual	2.8	1.7	
BodilyExp	Music	5.0	1.5	t (452.55) = 4.91; p ≤ .001
	Visual	4.2	1.8	

C. Correlations

Since *Aesthetic Experience*, *Attention* and *Bodily experience* demonstrated clear differences between the object types in emotional responsiveness, we further investigated whether they also showed connections to perceiving the object pleasurable. The object characteristic "Unpleasant-Pleasant", was therefore correlated with these emotional responsiveness features, separately in music (Table 4) and in visual objects (Table 5).

TABLE 4
OBJECT PLEASANTNESS & EMOTIONAL RESPONSIVENESS IN MUSIC

Variable	AesthExp	Attention	BodilyExp
Objec feature: Unpleasant-Pleasant	.16*	.00	.19**
AesthExp	1	.26**	.37**
Attention	.26**	.1	.33**
BodilyExp	.37**	.33**	1

* p ≤ .05; ** p ≤ .01; *** p ≤ .001

In music, the perceived pleasantness of the object was significantly correlated with the strength of Aesthetic experience and Bodily experience. For visual objects, the perceived pleasantness of the object correlated significantly only with the strength of Aesthetic experience. These correlations were relatively low but in line with earlier observations on modality differences in emotional reactivity towards the object.

TABLE 5
OBJECT PLEASANTNESS & EMOTIONAL RESPONSIVENESS IN VISUAL OBJECTS

Variable	AesthExp	Attention	BodilyExp
Objec feature: Unpleasant-Pleasant	.13*	.09	.01
AesthExp	1	.35**	.33**
Attention	.35**	.1	.32**
BodilyExp	.32**	.32**	1

* p ≤ .05; ** p ≤ .01; *** p ≤ .001

In regard to the components of emotional reactivity, Aesthetic experience, Attention, and Bodily experience, all were moderately positively correlated with each other, both for music and visual objects.

IV. DISCUSSION

The results provided evidence of distinct characteristics of pleasurable music and pleasurable visual objects, and the relation of these to emotional responsiveness. First, it was shown that both pleasurable music and pleasurable visual objects were perceived as *pleasant, timeless, nostalgic, sophisticated*, and *imposing,* on the bi-polar scales (as opposed to *unpleasant, temporal, oriented towards novelty, rough*, and *modest*, respectively).

In regards to differences between the modalities, it was found that pleasurable music was characterized more towards *serious, rebellious, mystical, bold, heavy, exciting, energetic, restless,* and *sensual*. In contrast, pleasurable visual objects were characterized more towards *playful* and *funny, conventional, down to earth, discreet, light, calming, calm, tranquil,* and *non-erotic.* These differences present pleasurable music and pleasurable visual objects as distinct, and appreciated in different ways. Music presents itself as a source of serious, mystical, heavy, and rebellious excitement, while visual objects offer playful, conventional, and light tranquility.

Results regarding emotional responsiveness provide an interesting additional perspective. Music received higher ratings in the strength of attention and bodily experience, and was generally emotionally reacted to more strongly than visual objects. Pleasure evoked by visual objects was experienced predominantly as aesthetic in nature. These differences were further supported by correlations between emotional responsiveness components and the object feature unpleasant-pleasant: Object pleasantness was only correlated with the strength of aesthetic response in visual objects but with both aesthetic and bodily response in music.

The limitations of the study concern sample biases in terms of gender, age, and education level, so one should be careful in making generalizations to other types of samples. Also, cultural environment may have influenced how emotions evoked by daily experiences of music and visual objects are perceived and conceptualized, and results may be somewhat reflective of Finnish culture.

Applying SDS is also somewhat problematic in measuring perceptions: because SDS does not include a built-in method of assessing reliability, the content validity of SDS is not as rigorous as in Likert scales due to increased ambiguity in concepts, and people with lower levels of education have shown to abandon the middle points of the scale and focus on the extremes (Sommer & Sommer, 1997). The final limitation was not observed in this study, probably due to the respondents' high education level. Nonetheless, it is always possible that the terms might have had slightly different interpretations between participants.

Regardless of these limitations, the study was able to provide valuable new knowledge to the research of music and visual culture. Our study established particular characteristics in perceptual features and emotional responsiveness, demonstrating distinct profiles of pleasure evoked by music and visual objects.

ACKNOWLEDGMENT

The study was funded by the Kone foundation.

REFERENCES

[1] D. R. Heise, *Surveying cultures: Discovering shared conceptions and sentiments*. Hoboken NJ: Wiley, 2010.

[2] S. Himmelfarb, "The measurement of attitudes," in *Psychology of Attitudes*, A. H. Eagly & S. Chaiken, Eds. Thomson/Wadsworth, 1993, pp. 23-88.

[3] P.W. Jordan, "*The pleasure of products,*" in *Pleasure with Products: Beyond Usability*, W. S. Green & P. W. Jordan, Eds. London: Taylor & Francis, 2002.

[4] C. E. Osgood, Suci, G. J., and Tannenbaum, P. H., *The measurement of meaning*, Urbana: University of Illinois Press, 1957.

[5] B. Sommer & Sommer, R., *A Practical guide to behavioral research: Tools and techniques*, New York: Oxford University Press.

APPENDIXES

Appendix 1: Semantic differential scale and mean values

Please estimate the features of the object of your selection according to the following adjective pairs. In your answer, please consider whether your experience corresponds strongly with either one of the given options (numbers 1-3 or 5-7) or settles neutrally between the given options (number 4).

Descriptive Statistics[a]

	Mean Music	Mean Visual
Pleasant - Unpleasant	1,487	1,534
Light - Heavy	3.724	3.021
Masculine - Feminine	4.132	4.449
Playful - Serious	4.342	3.466
Ordinary - Special	5.206	4.792
Ugly - Beautiful	6.434	6.428
Calming - Exciting	3.232	2.263
Temporal - Timeless	6.412	6.250
Intellectual - Emotional	5.224	4.792
Energetic - Calm	4.333	4.907
Conventional - Rebellious	4.154	3.644
Nostalgic - Oriented towards novelty	3.614	3.500
Cultural - Universal	4.684	4.521
Cold - Warm	5.904	5.475
Mystical - Down to earth	3.658	4.458
Vivid - Neutral	1.965	2.364
Complex - Simple	3.531	4.042
Predictable - Surprising	4.645	4.831
New - Old	4.246	4.432
Funny - Serious	4.754	3.992
Ambivalent - Harmonious	5.053	5.398
Trendy - Traditional	4.684	5.030
Luxurious - Ordinary	3.803	4.013
Natural - Technical	2.360	2.174
Tranquil - Restless	2.737	2.085
Thrilling - Secure	4.303	4.636
Soft, roundish - Hard, angular	2.829	2.911
Rough - Sophisticated	4.781	4.860
Discreet - Bold	4.474	3.873
Creative - Unimaginative	1.943	2.174
Modest - Imposing	4.553	4.576
Sensual - Non-erotic	3.974	4.699
Authentic - Commercial	2.316	2.047
Valid N (listwise)	228	236

a. 1. Object type = Music and visual object

Appendix 2: Questions measuring emotional responsiveness

How strongly do you usually react to things on an emotional level?
1 = I am calm and usually I do not react to things emotionally
7 = I react very strongly emotionally

Please estimate the strength of the emotions you attach to the selected object.
1 = Very weak
7 = Extremely strong

How strong is the feeling of pleasure that the selected object evokes?
1 = Very weak
7 = Extremely strong

In your opinion, can the emotional experience evoked by the object you selected be described as an aesthetic experience?
1 = Not at all
7 = Yes, the experience is highly aesthetic

When the object you selected is present, how strongly is your attention directed to it?
1 = My attention is directed to other things
7 = My attention is fully directed to the object

Do you attach ambivalent emotions to the object?
1 = Not at all
7 = Very strongly

To what extent can the pleasure evoked by the object you selected be described as a bodily experience?
1 = Not at all
7 = Pleasure is strongly bodily in nature

Brazilian children's emotional responses to Wagner's music

Patricia Vanzella[1,2], Paulo E. Andrade[3], Olga V.C.A. Andrade[3], E. Glenn Schellenberg[4]
[1]Department of Music, University of Brasília, Brasília, DF, Brazil
[2]Unit for Neuroscience and Complex Systems, Federal University of ABC, São Bernardo do Campo, SP, Brazil
[3]Department of Pedagogical Studies, School of Elementary and Secondary Education, Colégio Criativo, Marília, SP, Brazil
[4]Department of Psychology, University of Toronto Mississauga, Mississauga, ON, Canada

Music induces emotions and influences moods. Mode, tempo, loudness, and complexity have systematic associations with specific emotions. In previous research on children's sensitivity to emotional cues in music, however, real music has rarely been used. Our participants were middle-class children (7-10 years of age, n = 195), adult musicians (n = 57), and adult nonmusicians, (n = 51), recruited in Brazil. The task required participants to match emotion labels (e.g., happiness) with 1-min instrumental excerpts taken from Wagner's operas. Eight excerpts, from five operas, were selected so that they varied systematically in terms of arousal and valence, and so that two came from each quadrant of emotion space as defined by the circumplex model. Excerpts were presented to groups of participants over speakers. For each excerpt, listeners chose one emotion from a set of eight emotion words that varied identically in terms of arousal and valence (happy, elated, loving, peaceful, angry, fearful, sad, and distressed). Emoticons were presented with each emotion term to make the task easier for children. Responses were considered correct if the word and excerpt were in the same quadrant. Children's performance greatly exceeded chance levels and matched that of adult nonmusicians. There was also monotonic improvement as a function of grade level, and musicians outperformed nonmusicians and children. Group differences arose primarily from the ability to match words and music on the basis of valence cues. Sensitivity to arousal cues varied little across groups or grades. Thus, even 7-year-olds are capable of making consistent emotional evaluations of complex music. The results imply that such music could be used more frequently with young children to enrich their musical environment.

The relative contributions of composition and visual and auditory performance cues to emotion perception: Comparing piano and violin performances

Marc R. Thompson[1], Jonna K. Vuoskoski[2], Eric F. Clarke[2]
[1]University of Jyväskylä, Finland
[2]University of Oxford, UK

Visual and auditory performance cues, as well as structural features of the music, can all communicate emotional information to observers. However, little is known about the relative contributions of these components to emotions perceived in music, or whether the relative contributions change across different instruments. In two experiments, we systematically varied the emotion conveyed by the structural features of the music, as well as the emotional expression communicated by the performer. In Experiment 1, a pianist performed four short pieces (sad, happy, scary, and tender-sounding) with four different expressive intentions: sad, happy, angry, and deadpan. The body movements of the pianist were tracked using optical motion capture. Subsequently, 31 participants rated the emotions that they perceived in the performances in audio-only, video-only, and audiovisual rating conditions. In Experiment 2, we replicated the experimental design using violin performances (violin performed the melody from the same four short pieces) with 34 different participants. To investigate the possible influence of visual performance cues on emotion perception in the audiovisual rating condition, the four deadpan audio tracks were paired with motion capture animations representing the four different emotional expressions (of the same piece). A time-warping algorithm was applied to the motion capture animations so that the audio and video would appear synchronized. The ratings from each study were analyzed using repeated-measures ANOVAs. Our presentation focuses on how emotion ratings change depending on which instrument is being observed. For instance, in the video-only condition, participants viewing the piano stimuli accurately recognized the emotional expression conveyed by the performer, but those viewing violin stimuli confused emotional expressions. The varying results suggest that pianists and violinists convey emotional information in different ways to observers.

Building an expressive performance

Shawn Michael Condon
Department of Music, University of Jyväskylä, Jyväskylä, Finland

Emotional expressivity in performance requires singers to employ a careful balance of creativity, personal experience, and empathy. For audience members, the combination of these three elements can determine the overall success of a singer's performance. The significance of emotional expressivity in performance is generally accepted, but in educational settings is it often overlooked due to its intangible nature. This study examines the techniques used by vocalists when preparing a performance, from their first experiences with unfamiliar music and throughout the rehearsal process, until their final performance. Participants were enrolled in classical music programs at universities in the United States, Hong Kong, Finland, and Ireland. Survey responses were both qualitative and quantitative. Qualitative data was analyzed using thematic analysis in four categories: creating an expressive performance, defining expressivity, emotions in performance, and evidence of expressivity. The study identified individual's preferences for pedagogical methods and when specific expressive methods were used during the rehearsal process. During performance preparation, the performer is engaged two distinct phases; a learning phase characterized by methods for familiarization and building connections and a performance phase characterized by greater attention to their outward appearance, communicative expressivity, and emotional engagement. A clearer understanding of the development of emotional expressivity can inform pedagogical methodology and assist in more easily incorporating teaching styles that discuss emotional expressivity.

The role of perceived exertion in models of continuously perceived affect in response to music: Investigating the 'FEELA' hypothesis

Kirk N. Olsen[1], Roger T. Dean[1]
[1]The MARCS Institute for Brain, Behaviour, and Development, Western Sydney University, Sydney, Australia

Time-series modeling of perceived affect in response to a range of instrumental and sound based music has shown that continuously perceived arousal, and to a lesser extent valence, are well modeled when predictors include listener engagement, perceptual loudness, and acoustic factors such as intensity and spectral flatness. A 'FEELA' hypothesis has been proposed to explain processes underlying production and perception of affect in music: FEELA suggests a chain of contributing factors such as the Force and Effort (realized here throughout as physical exertion) required for a performer to produce a musical sound, the Energy of the resulting sound (realized as acoustic intensity), and the experience of the listener in the form of perceived Loudness and Arousal. The present study investigated the early portions of this process by asking whether listeners' continuous perception of physical exertion required to produce music from a range of genres contributes to and strengthens previous time-series models of perceived affect. An analysis of factors in the perception of exertion in the context of eight excerpts of Classical and Electroacoustic music was first undertaken. Results showed that acoustic intensity, perceived source complexity, and event density contributed in varying degrees to the perception of exertion. When all four factors were included in time-series models of affect, results showed that non-musicians' perception of exertion required to physically produce Classical or Electroacoustic music is pertinent to the perception of arousal, but only when human agency is apparent. With the more abstract sound-sculpted electroacoustic pieces of music where human agency is not always apparent, listeners could identify exertion when required, but it was not influential in their perception of arousal.

New integrative model of musical affect regulation

Margarida Baltazar[1], Suvi Saarikallio[1]
[1]Finnish Center for Interdisciplinary Music Research, Department of Music, University of Jyväskylä, Finland

Research on music uses has shown that regulation functions are transversely present across affect dimensions (e.g., emotions, energy levels, motivation) and populations. Nevertheless, a recent literature review (Baltazar & Saarikallio, submitted) revealed that no comprehensive model for the phenomenon exists and that music use is not equally studied across the affective dimensions. The current study aimed to empirically explore and clarify how contextual and individual factors, music features, affective goals, tactics, strategies, and mechanisms interact in the regulation process in order to develop a new model for musical affect regulation. Data were collected through an online survey, filled by 571 participants aged between 13 and 30. Preliminary results support a general model of musical affect regulation, as 34.7% reported a co-presence of two or more dimensions and 29.4% could not decide on one dimension. The most salient features of the process for the participants were strategies and mechanisms, chosen significantly more often as the motive for music engagement than goals or tactics. Most used strategies were *distraction*, *concentration*, and *manipulation of affective states*, and most common mechanisms were *rhythm*, *lyrics*, and *aesthetics*. However, the use of strategies and mechanisms varied depending on which of these two was more relevant for the participant. Detailed analyses regarding the co-occurrence and interrelation of these elements and the mediator role of individual-contextual factors are being conducted. The results indicate a process that, although based on the experienced state and individual-situational factors, is perceptually initiated with a focus on one or multiple regulatory strategies or mechanisms. The motivational focus influences which elements become part of the process. A new integrative model of affect regulation through music, combining the elements that emerged both from theoretical and empirical work, is proposed.

A comparison of machine learning methods as models for predicting music emotion appraisals

Naresh N. Vempala[1], Frank A. Russo[1]
[1] SMART lab, Department of Psychology, Ryerson University, Toronto, Ontario, Canada

In a previous study, we explored the classic debate on music emotion involving cognitivist and emotivist positions using connectionist neural networks. The cognitivist position states that music expresses emotion without inducing it in listeners. The emotivist position states that music induces emotions that resemble everyday emotions, enabling listeners to feel emotion, as supported by their physiological responses. Since the cognitivist position involves emotion appraisals of listeners based only on perception of emotion, we modeled this view using neural networks trained on audio features abstracted from music. If the emotivist position were true, listeners would feel emotion. Hence we modeled the emotivist view using listeners' physiological responses to music. We also assessed a third hypothesis, whereby listeners are assumed to base emotion appraisals on some combination of what they perceive in the music and experience through their own bodies. We modeled this hybrid position using committee machines. Despite promising results, the generalizability of our findings was limited due to the small size of our dataset consisting of only 12 classical music excerpts. Here, we revisit these models using a larger dataset of 60 excerpts, chosen from four different musical genres. We collected valence and arousal ratings of felt emotion, and physiological responses from 60 participants such that 12 unique listeners heard each excerpt. We present our revised findings in the context of various methodological improvements from our previous models that include feature reduction, cross-validation, and bagging to make the models more robust. Our modeling exercises reveal that arousal assessments are more driven by physiological features while valence assessments are more driven by audio features. Finally, we compare results from different models and highlight salient features for predicting emotion.

Imprinting Emotion on Music: Transferring Affective Information From Sight To Sound

Caitlyn Trevor,*1 Joseph Plazak#2

*Music, Ohio State University, United States of America
#Music, Illinois Wesleyan University, United States of America
[1]trevor.5@osu.edu, [2]jplazak@iwu.edu

ABSTRACT

In this study, we investigated the transferability of visual emotional information from film clips onto emotionally "neutral" musical works. We solicited short piano compositions that aimed to convey a sense of "neutrality," or at least the absence of strong emotions, from a pool of composers, and then paired each neutral composition with emotional film clips. Using a between subjects design, 20 participants provided subjective ratings of valence and arousal after watching these short film clips. For a given neutral composition, half of the participants heard the music paired with a nominally happy video clip, whereas the other half heard the same music with a nominally sad video clip. After a short distraction task, participants listened to the previously heard music samples, this time without the accompanying visual information and were asked to again provide subjective ratings of valence and arousal levels. We hypothesized that affective visual information would "imprint" onto neutral compositions, such that those who heard neutral music paired with a nominally happy film would rate the neutral music more positively than those who first heard the music paired with a nominally sad film. Our results investigated the reciprocal relationship between visual and aurally inferred emotional information, specifically by examining the influence of visual information on the aural experience.

I. INTRODUCTION

Much of the research on music and film is focused on how music affects the interpretation of the film (Cohen, 2001). For example, Marshall and Cohen (1998) showed that the perception of a film with animated geometric figures could be influenced by an accompanying musical sample. In their discussion, they outlined an explanatory model referred to as the congruency-associationist model (Figure 1) (Marshall & Cohen, 1988). In this figure, the Venn Circles represent the total meanings of the film and the music. When experiencing both the meaning of the film and the music, more attention is drawn towards characteristics that are the same (a). Information from the music itself (x) is then associated with the congruent characteristics of the film and music, thereby allowing the music to alter the meaning the film has when shown by itself.

While much of multimodal research demonstrates how aural information can influence the perception of visual information, other studies have shown that the opposite relationship also exists: visual information can influence aural experience.

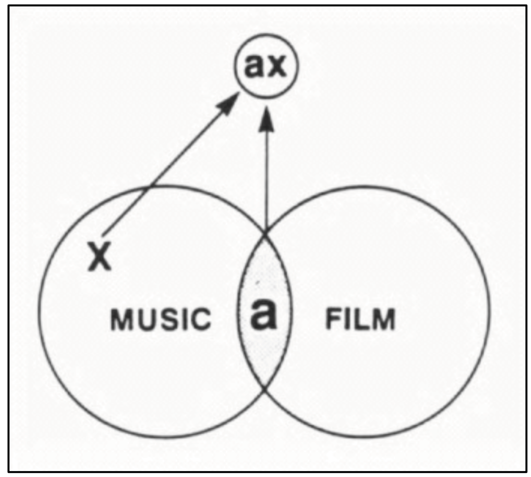


Figure 1. The Congruency-Associationist Model (Marshall& Cohen, 1988)

For example, McGurk and MacdDonald (1976) demonstrated that when participants were exposed to films of a young woman mouthing the syllable "ga" with a dubbed voice saying "ba", they reported hearing the syllable "da." The visual information from the woman's mouthing had altered the participants' perception of the aural information from the pronounced syllable. Another study demonstrated this phenomenon using filmed and recorded marimba tones. Schutz & Lipscomb (2007) showed that switching the video of a percussionist playing short and long marimba tones with the audio of these short and long tones could alter the perceived length of each tone. Both of these studies demonstrated that it is possible for visual information to alter the perception of aural information.

Given this possible "reversed relationship" between the senses, the question that this study explored was: can the congruency-associationist model also operate in reverse? Boltz, Ebendorf, & Field (2009) tested this conjecture by pairing positively and negatively valenced visuals with ambiguous musical samples and found that the visual information affected the perception of the musical samples in a mood-congruent manner. Specifically, ambiguous musical samples were rated more positively when paired with a positively affective visual and were rated more negatively when paired with a negatively affective visual.

Here we conducted a similar study by investigating the influence of positive or negatively valenced film clips on emotionally "neutral" music. Specifically, we predicted the following: *(H1) Affective information from silent-films will alter the perceived emotionality of "neutral" music.*

Specifically, affective information from short film clips will "bind" to relatively neutral music.

II. STUDY 1: EMOTIONALLY NEUTRAL MUSIC

A. Motivation

For our study design, it was necessary to acquire emotionally neutral music to accompany visual stimuli. The task of finding existing music that could be described as emotionally neutral proved to be difficult. Therefore, we elected to explicitly solicit emotionally "neutral" music from student composers. This solicitation presented a unique opportunity to study what musical characteristics might be perceived as emotionally neutral; an interesting question that will only be briefly discussed in this short report.

B. Participants

Six composition students from Illinois Wesleyan University between the ages of 18-22 were recruited to write emotionally "neutral" compositions.

C. Methodology

In an open call, composers were asked to write a short (no longer than 45 seconds) piano piece using standard music notation that expressed a sense of emotional neutrality. The composers were solely responsible for interpreting the meaning of "emotional neutrality." The instructions that they received were as follows: "*Composer participants will write a short piece (< 45s) for piano that they consider to be emotionally "neutral." Please interpret "neutral" however you see fit.*"

D. Analysis

Each piece was analysed with the Humdrum toolkit as well as other standard techniques. A brief descriptive analysis appears in Table 1.

E. Results and Discussion

The general characteristics of these pieces include an average length of 20.5 measures with a median of 4.12 notes per measure. The texture of these short pieces was generally fairly sparse with an average of 3 parts. The pitch range spanned just less than three octaves from B2 to Ab5; a range that falls comfortably within the average span of the human voice. In particular, this is an interesting finding given that more extreme ranges choices have been shown to yield a stronger emotional response, with higher range being rated as more tense and less stable (Costa, Bitti, & Bonfiglioli, 2000; Gabrielsson & Lindström, 2010). Within each piece, however, the median range was 25 semitones, just over two octaves.

Meters within our corpus of emotionally neutral music included duple and quadruple. With regards to metric regularity, we noted an absence of syncopation. Regarding tonality, it is interesting to note that the compositions tended to avoid major or minor thirds or sixths, both melodically and harmonically. Mode is often indicated by the quality of thirds and sixths, and further, mode is commonly associated with valence (Gabrielsson & Lindström, 2010). It is worth noting that some of the most common melodic and harmonic intervals, the P5, P4, P1, and P8 are described in music and emotion research as being more "carefree" (Maher & Berlyne, 1982; Gabrielsson & Lindström, 2010).

Table 1. Characteristics of the "neutral" compositions

Category	*Neutral Characteristics*
Average # of Measures	20.5 measures
Median # of Notes per Measure (all voices)	4.12 notes
Average # of Independent Parts	3 parts
Pitch Range (all pieces)	B2-Ab5
Median Range	25 semitones
Most Common Melodic Intervals	Unison, Octave, Major 2nd
Most Common Harmonic Intervals	Perfect 5th, Perfect 4th, Major 2nd
Meter	All compositions in duple or quadruple
Metric Regularity	Absence of syncopation

III. STUDY 2; METHODOLOGY

In our main study, we investigated whether film can leave a programmatic emotional impression on a piece of music that accompanies it. Below, the details of the main study are outlined.

F. Participants

Twenty participants were recruited from the Ohio State School of Music aural skills classes. These were primarily music majors between the ages of 18-22.

G. Materials

All neutral compositions collected from the first study were paired with short, 45-second, "valenced video clips." A total of eight video clips, taken from pre-existing online films, were selected according the following criterion. Each video was required to be clearly positively or negatively valenced, highly unfamiliar, a length of 45 seconds, and not involving overly graphic visuals, such as in the subjects of war or death. Audio and visual components were combined using iMovie.

H. Procedure

Participants were seated in front of a computer in a quiet room. Experimental software, designed using Max/MSP/Jitter, presented all of the experimental stimuli and collected all participant data (Cycling '74). The experimenter first familiarized each participant with the experiment software and allowed them to adjust the audio volume to a comfortable level. Care was taken to ensure that each participant understood the experimental interface (see Figure 3), the meaning of valance, arousal, and the circumplex model (Russell, 1980). Once participants were comfortable with the experiment interface, the experimenter read the following instructions:

"Six video clips will now be shown to you. Each video clip is approximately 45 seconds long. Between each clip, there will be a ten second pause. After the clips, you will be asked to rate the valence and arousal of each clip on the circumplex model you were previously shown. This will take approximately 7 minutes. Following this will be some paperwork about your basic information and demographics. There will then be a second stage in which you will be asked to listen to six audio files. After each audio file, you will be asked to again rate the valence and arousal on the circumplex model. The entire study will take approximately 20 minutes. If you become uncomfortable with any part of this study at any time, you may leave immediately."

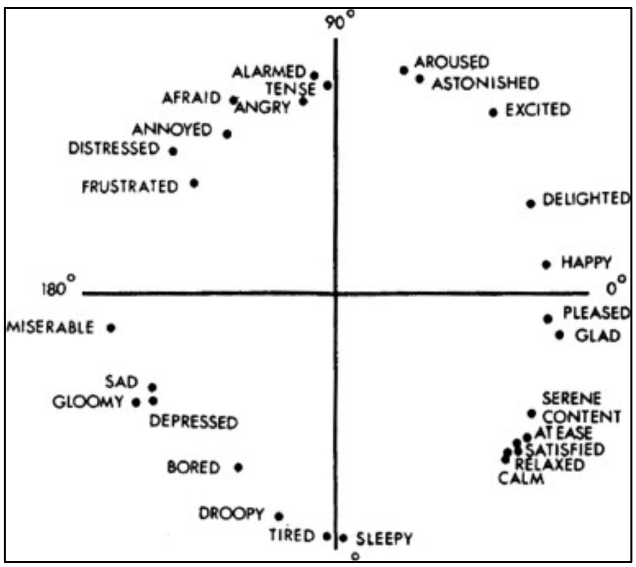


Figure 2. The Circumplex Model of Emotions (Russell, 1980).

Data collection took part in two stages. In the first stage, participants were asked to watch 6 short movie clips, and then rate valance and arousal levels via the circumplex model after each clip. After inputting their responses, participants clicked a button to initiate the next trial. Participants were only allowed to watch each video clip once. Stage 1 required approximately 7 minutes to complete. Participants completed a short distractor task between stage 1 and stage 2. Specifically, participants were asked to fill out a questionnaire consisting of basic demographic information. This distractor task took approximately 5 minutes. Stage two of the experiment was identical to stage 1, except that visual information was omitted. Participants heard the same 6 musical excerpts that were presented in stage 1, but this time without the accompanied "valenced videos". Participants were only allowed to listen to each audio file once, after which they again provided ratings of valence and arousal via the circumplex model (Figure 1). Stage 2 also required approximately 7 minutes to complete.

I. Experimental Design

Participants were randomly assigned to 1 of 2 experimental groups. All participants heard the 6 neutral compositions from study 1. However, groups were balanced so that when listening to any given neutral composition, half of the participants viewed a positively valenced video clip, whereas the other half of participants viewed a negatively valenced video clip.

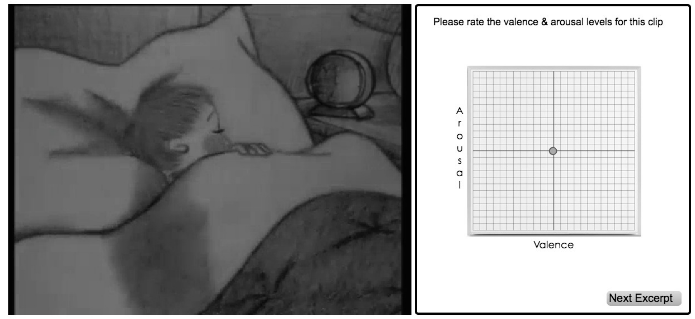


Figure 3. Screenshot of the experimental interface.

IV. RESULTS

Recall that our hypothesis predicted that the neutral compositions would be rated differently depending on which movie they were seen with due to the differences in valence and arousal between the movie clips. First it was necessary to verify that participants perceived different valence and arousal levels from the counterbalanced pairs of films. Therefore, t-tests for each pair of video clips paired with the same soundtrack were calculated. For the majority of the pairings, there was a significant difference in both the valence and arousal ratings (Table 1).

Table 1. The results for the valence ratings of the video and audio condition in which the ratings of the counterbalanced videos with the same audio track are compared (p-values were Bonferroni-corrected, α = 0.0083)

Valence Results of Video & Audio Condition

Track #	Emotion of Video Clip	Mean Valence	t-value	p-value
1	Happy	24.7	-4.3	0.0005*
	Sad	-46.5		
2	Happy	6.1	-3.67	0.004*
	Sad	-52.6		
3	Happy	-19.9	-3.13	0.008*
	Sad	-66.3		
4	Happy	-2.3	1.82	0.09
	Sad	-34.4		
5	Happy	72.6	5.89	0.00002*
	Sad	-38.3		
6	Happy	45.6	3.13	0.006*
	Sad	-26.1		

T-tests were also run for the arousal ratings of the video condition. There were no significant differences between the arousal ratings of the counterbalanced video pairs. With this difference established between the perceived valence of the film and audio pairs, we next tested our main hypothesis with *t*-tests to compare the valence and arousal ratings between the counterbalanced soundtracks in the audio-only condition. These results are presented in Tables 3 and 4.

Table 2. The results for the arousal ratings of the video and audio condition in which the ratings of the counterbalanced videos with the same audio track are compared (Bonferroni-corrected, α = 0.0083)

Arousal Results for Video & Audio Condition

Track #	Emotion of Video Clip	Mean Arousal	t-value	p-value
1	Happy	44.5	-2.79	0.014
	Sad	-18.4		
2	Happy	-50.1	-1.49	0.167
	Sad	-28.7		
3	Happy	10.4	-0.249	0.807
	Sad	4.9		
4	Happy	-14.9	0.315	0.757
	Sad	-21.3		
5	Happy	29.6	2.799	0.0128
	Sad	-30.5		
6	Happy	-1.8	2.799	0.015
	Sad	-61.9		

Table 3. The results comparing the counterbalanced valence ratings for the audio only condition

Valence Results for Audio Only Condition

Track #	Emotion of Video Clip	Mean Valence	t-value	p-value
1	Happy	-22.2	-0.449	0.659
	Sad	-11		
2	Happy	-16.8	-0.806	0.431
	Sad	-30.2		
3	Happy	-43.3	-0.251	0.804
	Sad	-46.5		
4	Happy	-19.1	0.979	0.341
	Sad	-31.7		
5	Happy	15.8	0.555	0.586
	Sad	2.9		
6	Happy	7.5	-0.585	0.567
	Sad	20.3		

Table 4. The results comparing the counterbalanced arousal ratings for the audio only condition

Arousal Results for Audio Only Condition

Track #	Emotion of Video Clip	Mean Arousal	t-value	p-value
1	Happy	53	-0.04	0.969
	Sad	53		
2	Happy	-20.1	-0.746	0.467
	Sad	-8.4		
3	Happy	43.5	0.988	0.337
	Sad	25.9		
4	Happy	9.4	-0.559	0.583
	Sad	19.4		
5	Happy	-62.1	-1.149	0.269
	Sad	-39.3		
6	Happy	-71.6	-1.505	0.155
	Sad	-44.8		

As can be seen in the tables, none of the ratings of the audio-only counterbalanced soundtracks were significantly different from each other. Therefore, our results did not provide support for our hypothesis.

V. DISCUSSION

Despite significant differences between the valence and arousal ratings of the counterbalanced audio-visual pairs, our results did not find evidence that affective visual information could be "imprinted" onto neutral musical compositions (as predicted by H1). One plausible explanation for our results might be that the congruency-associationist model simply might not operate in reverse. However, we would like to propose an alternative interpretation of the data.

The congruency portion of the congruency-associationist model assumes similarity between the auditory and the visual domains (however small or large) in order to emphasize a shared meaning. Within our debriefing sessions, multiple subjects commented on the stark contrasts between the presented visual and auditory information. This leads us to believe that the congruency-associationist model failed to operate as we predicted partially due to a lack of audio-visual congruency in our stimuli. Future iterations of this research design might elect to deliberately align various musical elements (e.g., time rhythmic elements) of neutral audio tracks so as to fit (at least minimally) within the visual temporality of each film clip, thereby ensuring at least one congruent element. We believe that such an implementation could strengthen the congruency portion of our paradigm and potentially produce different results.

ACKNOWLEDGMENT

We would like to thank the Cognitive and Systematic Musicology Laboratory at Ohio State University for their feedback and support. We would also like to thank the Music Cognition Lab at Illinois Wesleyan University.

REFERENCES

Boltz, M. G., Ebendorf, B., & Field, B. (2009). Audiovisual interactions: The impact of visual information on music perception and memory. *Music Perception: An Interdisciplinary Journal*, *27*(1), 43-59.

Cohen, A. J. (2001). Music as a source of emotion in film. *Music and emotion: Theory and research, ed. PN Juslin & JA Sloboda*, 249-72.

Costa, M., Bitti, P. E. R., & Bonfiglioli, L. (2000). Psychological connotations of harmonic musical intervals. *Psychology of Music*, *28*(1), 4-22.

Gabrielsson, A., & Lindström, E. (2010). The role of structure in the musical expression of emotions. *Handbook of music and emotion: Theory, research, applications*, 367-400.

Maher, T. F., & Berlyne, D. E. (1982). Verbal and exploratory responses to melodic musical intervals. *Psychology of Music*.

Marshall, S. K., & Cohen, A. J. (1988). Effects of musical soundtracks on attitudes toward animated geometric figures. *Music Perception: An Interdisciplinary Journal*, *6*(1), 95-112.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746-748.

Russell, J. A. (1980). A Circumplex Model of Affect. *Journal of Personality and Social PsychologyI,* 39, 1161 – 1178.

Schutz, M., & Lipscomb, S. (2007). Hearing gestures, seeing music: Vision influences perceived tone duration. *Perception*, *36*(6), 888-897.

Matching Music to Brand Personality: A Semantic Differential Tool for Measuring Emotional Space

David John Baker,*[1] Tabitha Trahan,[2#] Daniel Müllensiefen[3!]

*Louisiana State University School of Music, Louisiana State University, USA
#Soundout, London, England
!Goldsmiths, University of London, London, England

[1]davidjohnbaker1@gmail.com, [2]tabi@soundout.com, [3]d.mullensiefen@gold.ac.uk

ABSTRACT

Advertising and branding practitioners in music industry are often faced with the task of finding and selecting music that matches a given brand profile to enhance the overall brand perception or impact of a commercial. Currently there are only few suggestions in the literature on how to best match music to brands (Brodsky, 2011; MacInnins & Park, 1991) and these largely lack the ability to quantify the brand-music relationships precisely. In order to create a practical tool for use in industry, a short psychometric tool was constructed to accurately quantify the perceived distance between brands and musical pieces in an emotional space. The tool combines research from music cognition (Asmus, 1985), music and advertising (Müllensiefen et. al 2013), as well as input from industry professionals. The semantic differential tool was created using previous data where 185 participants made 700 ratings of perceived affect for 16 music pieces across the 39 items representing a 3-factor structure suggested by Asmus (1985). On the basis of a factor analysis a set of 15 items was identified that represented the 3 emotional factors adequately. In a second step, new data was collected using these 15 items across 60 pieces of music presented via Soundout's SliceThePie.com interface (N=6005). A confirmatory factor analysis confirmed the 3-factor structure of the tool within acceptable bounds (RMR <.08, RMSEA <.101,) with factors being dubbed Dark, Vibrant, and Tranquil (Cronbach's alpha = 88, .85, .87 respectively). Ratings of music by a brand's target audience can now be compared to brand ratings by industry professionals using Euclidian distances, resulting in a distance score between song and brand. This measure has been implemented by SoundOut as part for their SyncSight research toolbox and allows for fast, quantitative comparisons that can inform choices regarding the selection of music to appropriately fit brand personalities or suit advertising campaigns.

I. INTRODUCTION

Advertising and branding practitioners in music industry are often faced with the task of finding and selecting music that matches a given brand profile to enhance the overall brand perception or impact of a commercial. Most often than not, this task is left up to the 'Creatives' on an advertising team. Using their experience and expert opinion a Creative will work with a client and propose suggestions based on the client's needs. Though it is at this point that problems arise since the clients often may not be able to accurately articulate what they need. More often than not clients will make vague requests and may not even have a comprehensive understanding of their own brand personality. It is then the task of the Creative to come up with a list of music options to satisfy the needs of their clients. Situations like these are rife with opportunities for miscommunication that can lead to frustration on both sides.

Currently there are only a few suggestions on how to best select music for brands. Early research on systematizing the process can be found from MacInnins and Park (1991) who introduced the concept of fitting brands to music that were taken up by later researchers. Brodsky (2011) uses quantitative techniques to demonstrate how careful modeling can help more accurately classify brand fit within two car branding options, but does not offer any sort of generalizable methodology or specific tool for other researchers to use. Despite these efforts, the research largely lacks the ability to quantify brand-music relationships, or fit, accurately. In order to create a practical instrument for use in industry, a short psychometric tool was constructed to accurately quantify the perceived distance between brands and musical pieces in a emotional space: a concept that we see analogous to the idea of brand-music fit. This tool combines research on music cognition (Asmus, 1985), music and advertising (Müllenseifen et. al, 2013), as well as input from industry professionals. In this paper we present our methodology in creating a tool to best classify brand to an emotional space.

II. Methods

This tool was developed by first performing an exploratory factor analysis (EFA) on previously collected music and advertising data. The factor structure was then confirmed using confirmatory factory analysis techniques (CFA). The tool provides factor scores to serve as the basis for two separate measures for song as well as brand profile which can be compared using the Euclidian distance measure.

A. Exploratory Factor Analysis

The first stage of research extended previous work on music and advertising done in conjunction with the Adam& EveDDB advertising firm based in London, England. Data from Müllenseifen et. al (2013) was used that collected 700 individual ratings from 185 different participants of perceived affect for 16 music pieces across the 39 items representing the 3 factor structure suggested by Asmus (1985). The data contains ratings made by participants over a web study in which they were asked to "Please rate this track as if it were a person, to what extent would you describe them on each dimension" on a seven point Likert scale. Each participant rated four songs across the 39 items; a detailed report of the methodology can be found in Müllensiefen et al. (2013). This data was then used as the basis for the initial EFA. A listing of the initial 39 items can be found in Table 1.

Table 1. Table of 39 Initial Attributes used in Study

Victorious	Angry	Peaceful	Heroic
Raging	Calm	Stately	Cruel
Relaxed	Majestic	Hateful	Gentle
Determined	Frustrated	Pleasant	Vibrant
Depressive	Contemplative	Vigorous	Dreary
Reflective	Exuberant	Blue	Serene
Comical	Sad	Tranquil	Humorous
Gloomy	Loving	Amusing	Yearning
Tender	Playful	Longing	Beautiful
Cheerful	Lonely	Romantic	

B. Factor Analysis and Removal of Items

A exploratory factor analysis with oblimin rotation was run on the data. After the initial factor analysis, we then sought to reduce the number of items from 39 to 15. Our first step in the removal of items was to remove items that did not seem to effectively describe each factor as a whole. Here we removed the adjectives victorious, heroic, stately, majestic, determined, angry, raging, cruel, hateful, and pleasant. After removing these items, we then looked at each item's factor score across all three factors and removed items that did not seem to load clearly on a single factor, which we operationalized as having a loading of less than .5 on all factors .This excluded cheerful, longing, and contemplative.

Next we eliminated items that scored high on item uniqueness thus excluding vigorous, frustrated, related, loving, and beautiful. Our last exclusion criterion to reduce the amount of items was to remove items that appeared to be synonyms with other included items, which then removed the word reflective since two native English speakers deemed it synonymous with contemplative. The final subset of items retained along with their factor loadings can be seen in Table 2. The three factors were interpreted as Vibrant, Dark, and Tranquil in collaboration with industry professionals that were going to implement the tool.

Table 2. Factors, Items, and Their Loadings from EFA Model

FACTOR	LOADING	ITEM
Vibrant	Vibrant	.623
	Exuberant	.635
	Humorous	.956
	Amusing	.954
	Playful	.711
Dark	Depressive	.891
	Blue	.727
	Sad	.941
	Lonely	.768
Tranquil	Peaceful	.903
	Calm	.837
	Gentle	.835
	Serene	.783
	Tranquil	.867
	Tender	.662

C. Confirmatory Factor Analysis

After the EFA, the 15 items were then presented to participants in a web-study via Soundout's SliceThePie.com interface and N=6006 ratings over 60 songs (each participant rated one song) were collected. Participants were asked the same prompt in the original study "Please rate this track as if it were a person, to what extent would you describe them on each dimension?" but instead responded on a ten-point scale. Participants also made ratings on a liking and a similarity prompt that is to be used in future tool creation. The similarity and liking data was not used in the confirmatory factor analysis.

D. Confirmatory Factor Analysis Model Fitting

The initial confirmatory factor analysis accounted for all 15 items on their respective factors. While some of the fit indices in Model 1 are acceptable in terms of confirming the initial hypothesis (Standardized Root Mean Square Residual SRMR <.08), the Root Mean Square Error of Approximation (RMSEA) did not reach an acceptable threshold of <.1. This could be due to the fact that modifications were made post hoc to the initial EFA in order to accommodate industry concerns. In order to investigate this, we removed five items (comical, victorious, relaxed, loving, romantic) from the CFA that were only present in the initial EFA. Removal of these items did decrease the RMSEA from .102 to .085, which has been deemed an acceptable level for the CFA fit. Additionally, the problem of fit was also explored by reducing the amount of items in the Tranquil factor due to a non-significant relationship between Vibrancy and Tranquil in the first model. A third analysis with a reduced number of items from each factor was created. By dropping some of the items that did not load as highly on the initial EFA, the RMSEA has now dropped again to .071. Additionally, this final model with 6 fewer items from the first model eliminated the non-significant covariance between the Vibrancy and Tranquil factors. The three resulting factors yielded Cronbach's alpha scores of .88 for the Vibrancy factor, .85 for the Dark factor, and .87 for the Tranquil factor.

E. Calculation of Euclidian Distance to From Semantic Differential

Using the item ratings from Table 3, it is then possible to create a semantic differential relationship between any two items measured with the tool. If a client's goal is to match music to brand in order to calculate brand fit, a brand manager would then to need to rate his or her perception of their brand using the same prompt used for the music. We highly recommend that multiple brand managers make these ratings independently of one another and check for inter rater reliability.

After this process is complete, each rating can then be multiplied by the factor scores shown in Table 3 which then creates a metric that the songs can be objectively compared to. This same process is done for each song that was rated. This standardized metric can then be used to calculate the distance measure on each of the three factors shown in in Table 3 or a combined metric can be computed taking all three factors in the table into account simultaneously. An example of this can be seen in Figure 1. In Figure 1 each line represents the distance in two dimensional Euclidian space that either a song or brand exists in relation to one another. If a client is looking to match song and brand they will look for lines that appear to match up with one another. This process an be done quantitatively by also looking at the distance measures. If a line extends in a different dimension from the target, that choice should be avoided since it does not match accordingly.

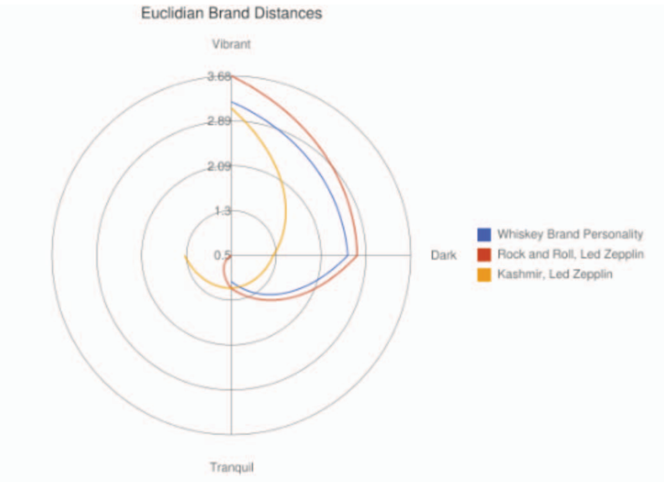


Figure 1 Example of Mapping Euclidian Distance

III. DISCUSSION

Unlike tools created in the past to examine the relationship between music and brand, this new tool can now provide an objective reference point to frame discussions regarding the relationship between music and brand. The tool has been dubbed the SightSync and is used in a suite of music market research tools that is now implemented by the two companies involved in the creation of the tool: Soundout (https://www.soundout.com/) and Hush (http://www.hushmusic.tv/). The tool is easy to use and does not require any sort of commercial statistical software to run. By simply entering the factor values into most spreadsheet management software, ratings then can be compared between brand personality and music.

Beyond providing a starting point for conversations between brand managers and creatives, the tool can also serve as a perceptual similarity tool to find similarity between tracks of music that share the same affective categorization but may not share the same features such as copyright detailing.

This would be helpful for an advertising company looking to use a certain track for their campaign or ad, but may be limited in terms of funding for rights and royalties. If the advertising company knew they wanted an effect similar to a song that may be out of their price range, they could simply rate that song according to the SightSync tool, then use those scores to find similar songs from a larger database that scored closely to the target song. This would then allow the marketing company to hit their target affect without having to pay expensive licensing fees.

The development and creation of this tool also created interest in future directions and refinements of our current model. The largest critique from industry demands was that the language used from the original collection of items was dated and did not reflect current industry terminology. Industry professionals noted that a similar tool that measured value judgments of music as opposed to perceived emotion would be helpful.

The a second version of the tool is planned in the future that will follow similar methodology, but instead draw from more industry oversight in the creation of the original list of termed that would relate to different facets of value judgments of music, which would then be analyzed further suing more sophisticated techniques of language processing such as Latent Semantic Analysis or Natural Language Processing to get a more accurate understanding of factors used to group the music.

REFERENCES

Asmus, E. P. (1985). The development of a multidimensional instrument for the measurement of affective response to music. In P. *Psychology of Music, 13*(1), 19-30.

Brodsky, W. (2011). Developing a functional method to apply music in branding: Designing language-generated music. *Psychology of Music,* 0305735610387778.

MacInnis, D. J. & Park, C. W. (1991). The differential role of characteristic of music on high and low involvement of consumers' processing of ads. *Journal of Consumer Research,* 161-173.

Müllenseifen, D., Davies, C., Dossman, L., Hansen, J.L., and Pickering, A. (2013) Implicit and Explicit Effect of Music on Brand Perception in TV Ads. In K. Bronner, R. Hirt & C. Ringe (Eds.), *Audio Branding Academy Yearbook 2012/2013* (pp. 139-153). Baden-Baden, Germany: Nomos.

Audience Emotional Response to Live and Recorded Musical Performance

Elizabeth Monzingo

Department of Music Theory, University of Oklahoma, USA
emonzingo@ou.edu

ABSTRACT

Collegiate music curricula commonly require performance attendance, often for zero credit hours. The relatively recent increase in the accessibility of a broad range of music through technological means provides an opportunity for music students to "attend" recorded performances in place of live performances, which would allow students more flexibility in fulfilling their performance attendance requirements. Thus, a need exists for research into the value of attending a live performance rather than "attending" a recorded performance. The purpose of this study was to investigate the difference between attending a live or recorded performance through a cognitive approach, assessing the emotional responses of listeners. Seventy-nine participants were randomly assigned to one of four conditions: visual and aural access to a live piano performance, visual and aural access to a recording of that performance, solely aural access to the live performance, or solely aural access to the recorded performance. Participants also completed a musical sophistication index. Every thirty seconds during the performance, the participants were instructed to indicate their present emotional response via two five-point Likert scales, one to measure the level to which they found the music pleasant, and the other to measure the level to which they were emotionally aroused by the music. Statistically significant interactions ($p < .05$) were found between the effects of performance setting and visual access on three pleasantness scale responses during two pieces of music, suggesting live performance audiences preferred visual access to the performance while recorded performance audiences preferred no visual access to the performance. Simple main effects analysis revealed participants in the live performance conditions found the performance more pleasant on two pleasantness scale responses during one piece of music ($p < .04$) and more emotionally arousing on three arousal scale responses during two pieces of music ($p < .05$) than the participants in the recorded performance conditions. While digitizing musical performances increases accessibility to broad ranges of music, students are best served by attending live performances.

I. INTRODUCTION

Studies into the advantages of communication via in-person or recorded methods have interested researchers for some time, particularly with the continually improving technologies for immediate communication across vast distances. Much research supports the superiority of a live setting. Kuhl, Tsao, and Lieu (2003) found infants demonstrated phonetic learning from live exposure to a foreign language, but not from prerecorded foreign language exposure. Cochrane (2011) discovered live music caused an overall reduction in hospice patients' perceived pain, while recorded music did not have any significant effect on the perceived pain. This research suggests there may be an element of the communication between individuals that is lost when one party is not physically present.

Justin and Vastfjall (2008) present a different approach to understanding the interaction people have with music, studying it not as a communication but rather through six mechanisms through which music might induce emotions. Some of these mechanisms, such as one referred to as the "brain stem reflex," invoke primal responses theoretically resultant from evolutionary advances. Others rely on individual and contextual factors, such as when music evokes a memory or particular settings create specific expectations.

Another factor that may greatly influence an individual's response to a musical performance is their level of musicianship and their previous experience studying music. Adams (1994) studied the effect of visual and aural stimuli on musician's and non-musician's emotional responses induced by a recording of an opera. His results suggest more engaged listeners are more likely to indicate more interest and a larger range of emotional responses, and that musicians are more likely to be engaged by performances.

Previous research suggests that a more engaged audience will be more emotionally involved in the performance, and that a live context increases engagement. Musically sophisticated individuals also tend to be more engaged by musical performance than less musically sophisticated individuals. The present study investigates an audience's emotional response to a live or recorded musical performance, also altering the aural and visual access to the performance and testing for differences due to audience member musical sophistication. It is hypothesized that the audiences attending the live performance will have stronger emotional responses than the audiences attending the recorded performance, and the audiences with the ability to see and hear the performer will have stronger emotional responses than the audiences only able to hear the performer. Additionally, it is hypothesized that there will be a positive correlation between strong emotional response and musical experience of the audience member.

II. METHOD

A. Design

Participants were randomly assigned to one of four conditions, described in the Table 1. Participants in conditions 1 and 2 attended a single live performance in a university recital hall. Participants in condition 1 were seated at the back of the recital hall, and participants in condition 2 were seated at the back of the second level judges' balcony, from where they were unable to see the stage. Participants in conditions 3 and 4 attended a screening of the recording of the live performance, shown on a projector in the same recital hall, with participants in condition 3 sitting in the back of the recital hall and participants in condition 4 sitting in the judges' balcony.

Table 1. Experimental conditions.

	Audio and visual access to performance	**Audio access, no visual access to performance**
Live performance	Condition 1	Condition 2
Recorded performance	Condition 3	Condition 4

B. Stimuli

A Master of Music student performed roughly ten minutes of piano music, including: *Piano Sonata no. 32 in C op. 11, I*, by Beethoven; *Anamorfosi*, by Sciarrino; and *Liza*, by Gershwin and Wild. Every thirty seconds, a researcher displayed a prompt to the participants to indicate their present emotional response via two five-point Likert scales, found in Figures 1 and 2.

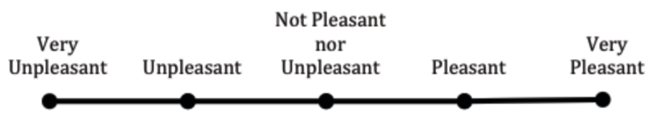


Figure 1. Pleasantness scale.

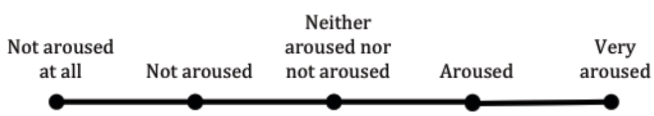


Figure 2. Arousal Scale.

C. Procedure

In each condition, participants first completed the Ollen Musical Sophistication Index on paper. The researcher then gave participants the following prepared instructions for the experiment:

"You will hear a musical performance. At various moments during the performance, you will see a numbered cue card reading, 'Please record your current emotional response.' When you see the cue, you will find its corresponding number on your response sheet and circle the phrase that best describes your current emotional response on both scales given. The musical performance will not pause for you to respond; respond quickly and continue to listen to the musical performance. The first scale asks you to record your emotional response on a scale from 'Very Unpleasant' to 'Very Pleasant.' The second scale asks you to record your emotional response on a scale from 'Not Aroused at All' to 'Very Aroused.' Arousal is a state of heightened physiological activity. For example, if you were to feel slightly guilty, you might assess your emotion as 'Unpleasant' as well as 'Aroused,' whereas if you were to feel very guilty, you might assess your emotion as 'Very Unpleasant' as well as 'Very Aroused.' For another example, consider the emotion, 'sleepy.' You might assess this emotion as 'Pleasant' as well as 'Not Aroused at All.'"

As the pianist played, a researcher held up a cue card every thirty seconds, prompting the participants to indicate their present emotional response on their response sheets.

D. Participants

79 university students (37 females and 42 males, ages ranging from 18 to 40 [M = 20]) participated in this experiment.

III. RESULTS

A two-way ANOVA was conducted to examine the effect of performance setting—live or recorded performance—and visual access—audio and visual access to the performance or only audio access to the performance—on responses on both the pleasantness scale and the emotional arousal scale. There was a statistically significant interaction between the effects of performance setting and visual access on pleasantness scale responses one minute into the Beethoven piece, F (2, 78) = 4.777, p = .032; 2.5 minutes into the Beethoven piece, F (2, 78) = 4.886, p = .030; and one minute into the Sciarrino piece, F (2, 78) = 4.010, p = .049. Interactions are displayed in Figures 3, 4, and 5.

Participants attending the live performance indicated significantly higher pleasantness at 2 minutes into the Gershwin/Wild piece, F (1, 75) = 4.756, p = .032, and 3.5 minutes into the same piece, F (1, 75) = 4.860, p = .031. Participants attending the live performance also indicated significantly higher emotional arousal at 30 seconds into the Beethoven piece, F (1, 75) = 4.141, p = .045; 2.5 minutes into the Gershwin/Wild piece, F (1, 75) = 13.377, p < .001; and 3 minutes into the Gershwin/Wild piece, F (1, 75) = 4.677, p = .034.

Musical sophistication was not found to significantly effect responses on either pleasantness nor emotional arousal scales in any condition.

Figure 3. Interaction between the effects of performance setting and visual access on pleasantness scale response 2, F (1, 75) = 4.777, p = .032.

Figure 4. Interaction between the effects of performance setting and visual access on pleasantness scale response 5, *F* (1, 75) = 4.886, *p* = .030.

Figure 5. Interaction between the effects of performance setting and visual access on pleasantness scale response 9, *F* (1, 75) = 4.010, *p* = .049.

IV. CONCLUSION AND DISCUSSION

Participants attending the live musical performance indicated higher levels of pleasantness and emotional arousal than did participants attending the recorded musical performance. The effect of visual access to the performance interacted with that of the performance setting, suggesting live performance audiences prefer to have visual access to the performance while recorded performance audiences prefer not to have visual access to the performance. There were no significant differences in pleasantness or arousal scale responses due to musical sophistication level. Significant differences were more likely to occur in the beginning or between the middle and the end of the individual pieces of music.

While access to musical performances is greatly increased by the technological advances in recording and sharing music, this study suggests students are best served by attending live performances as they are more likely to find the performance pleasant and emotionally arousing within a live performance setting than in a recorded performance setting. Preference for visual access in exclusively the live setting suggests the physical presence of the performer may impact audience members' desire to see the performer. Future research may benefit from studying the effect that psychological and physical closeness to the performer may have on audience emotional response to the performance.

REFERENCES

Adams, B. L. (1994). The effect of visual/aural conditions on the emotional response to music. (Doctoral Dissertation). Retrieved from ProQuest Dissertations and Theses. (Accession Order No. 9434127).

Cochrane, E. A. (2011). The effect of live versus recorded preferred sedative music on pain intensity perception in hospice patients. (Doctoral Dissertation). Retrieved from ProQuest Dissertations and Theses. (Accession Order No. 1494616).

Kuhl, P. K., Tsao F., and Liu, H. (2003). Foreign-Language Experience in Infancy: Effects of Short-Term Exposure and Social Interaction on Phonetic Learning. Proceedings of the National Academy of Sciences of the United States of America, Vol. 100, No. 15, pp. 9096-9101.

Vaughan, B. A. (1983). Comparison of the effects of recorded and live piano performance in listening experiences for sixth grade students. (Doctoral Dissertation). Retrieved from ProQuest Dissertations and Theses. (Accession Order No. 8322479).

The relationship between coupled oscillations during auditory temporal prediction and motor trajectories in a rhythmic synchronization task

Fiona C. Manning[1], Brandon T. Paul[1] and Serena Tse[1]
[1]Department of Psychology, Neuroscience and Behaviour, McMaster University, Canada

Our ability to predict upcoming events in music is thought to arise from a coupling of auditory and motor brain networks that establish expectations for when events (i.e., rhythmic sounds) are likely to occur. Evidence for this view is derived from neuroimaging studies showing engagement of low- (e.g., delta) and high- (e.g., beta) frequency oscillations sourced to auditory and motor regions that are time locked to predictable auditory events, when no overt movement is made by the listener. Despite these findings no direct relationship has been established between neural correlates of audio-motor coupling during auditory prediction and motor trajectory phases. Here we test for this relationship. Musicians and nonmusicians (right-handed) listened to isochronous tone sequences at two tempos while we recorded EEG and motion capture data simultaneously in separate phases of listening and tapping. When listening, participants identified deviant (louder) tones and did not synchronize movement to the sequence. In the tapping task, participants synchronized right index finger tapping with the sequence. Data collection and analysis are ongoing. Our analysis will correlate the strength of delta-beta coupling localized to auditory and motor regions in the listening task to consistency of movement trajectory phases. In a second analysis, we will compute the amount of mutual information shared between time series of neural activity during the listening task, neural activity during the tapping task, and motion trajectory time series during tapping. If we find high correspondence between auditory prediction networks and those involved in sensorimotor synchronization, we suggest auditory-motor linkages serve as a substrate for both rhythmically timed prediction and action, consistent with current views. If weak correspondence is found, we suggest an alternative view wherein auditory temporal prediction networks are complementary to those involved in timed actions.

Inducing Emotion Through Size-Related Timbre Manipulations: A Pilot Investigation

Joseph Plazak,*[1] Zachary Silver[2]

*Music Theory, Illinois Wesleyan University, United States of America
[1]jplazak@iwu.edu, [2]zsilver@iwu.edu

ABSTRACT

The ability to "hear" the size of a sound source (e.g., the size of an animal, the length of a dropped metal rod, the height of a human, etc.) has been well documented in psychoacoustic literature. Here, we present results from a pilot experiment that aimed to test if size-related timbre manipulations (i.e., spectral envelope modifications) could reliably explain "induced" emotional responses to music. Using a within-subject design, our paradigm first recorded participants singing a sustained musical tone in the middle portion of their singing range. Thereafter, we used the TANDEM-STRAIGHT vocoder to morph the spectral envelope of each recording so as to create the impression of tones produced by a singer that was either larger or smaller than the participant, thus uniquely normalizing the experimental stimuli to the size of the listener. Manipulated tones were presented in a counterbalanced fashion while measurements of physiological arousal (i.e., skin conductance) were simultaneously recorded. Participants also provided subjective ratings on the perceived sound source size of each tone. We hypothesized that listeners would be capable of classifying sounds as being larger or smaller than themselves, and more importantly, that relatively larger sounds would be more likely to result in induced emotional responses. The results from this study provide a novel insight for understanding induced musical affect by "normalizing" sound stimuli to each specific listener. This approach, combined with a better understanding of how sound source size is inferred for instrumental timbres, might reveal how various orchestrations, from sparse to grandiose, can reliably induce certain types of emotional responses via music.

I. INTRODUCTION

Research on the distinction between perceived and felt emotional communication in music has been well documented in music psychology literature (Gabrielsson, 2002; Schubert, 2013). For a variety of reasons, the majority of published literature on musical emotions tends to investigate perceived emotional responses, perhaps because the process of introspecting about a perceived emotional message may be easier than interpreting one's emotional state. On the other hand, "felt" emotional responses to music tend to be difficult to objectively measure in a laboratory setting, and often involve a great deal of inter-subject variability. The primary objective of our study was to test the utility of a novel theoretical concept (listener-normalized emotional responses) towards explaining inter-subject variability for felt emotional responses to music.

The main idea behind listener-normalized felt emotions (Plazak, in press) is that perceived musical emotions are normalized relative to the sound source, whereas felt emotions are normalized relative to the musical receiver. Listener-normalization relies on an ecological approach to music perception (Gaver, 1993; Clarke, 2006) and psychomechanical descriptors of sound (e.g. size, effort, proximity, etc.). In this particular study, we attempted to induce emotional responses in participants by manipulating size-related timbral cues that were normalized to the size of each participant. Using a high-quality vocoder (TANDEM-STRAIGHT), we were able to morph the spectral envelopes of recordings made by each participant so as to give the impression that the recording had been made by a sound source that was either larger or smaller (Kawahara et al., 2009). By normalizing stimuli relative to the sound source size of each listener, our aim was to reduce the inter-subject variability associated with physiological measurements of induced musical emotion within a controlled experiment, and to test if larger perceived sound sources are more likely to induce emotional responses.

Of those studies interested in "induced" or "felt" emotional responses to music, several methodological tools are available, including: physiological measurements, self-report experimental studies, journal studies, and qualitative in-depth interviews (Juslin & Laukka, 2004). Our primary focus in this study pertained to "induced" emotional responses to music recorded via physiological measurements of skin conductance. Skin conductance measurements are widely used as an indicator of physiological arousal (Boucsein, 2012), and the use of skin conductance to measure induced emotional experiences has been applied in several musical studies. According to Rickard (2004), participant-selected emotionally powerful music is particularly likely to elicit increases in skin conductance. Further, a study by Olsen & Stephens (2013) found that physiological arousal in response to basic attributes of music, in this case musical crescendi and decrescendi, can be differentiated via measurements of event-related skin conductance magnitude and skin conductance response rise times.

Our study sought to normalize the perceived sound source size of a sung musical tone relative to each participant. We hypothesized that listeners would be capable of classifying sounds as being larger or smaller than themselves, and more importantly, that relatively larger sounds would be more likely to result in induced emotional responses. The results from this study provide a novel insight for understanding induced musical affect through the lens of listener-normalized emotional responses.

II. METHOD

A. Participants

29 musician participants were recruited via email for the study. Participants were offered extra credit in a music theory course in exchange for their participation. All participants provided informed consent, and the Illinois Wesleyan

University Institutional Review Board certified this study for compliance with ethical standards.

B. Apparatus/Materials

Skin conductance measurements were collected by attaching two small gold-plated brass sensors (NeuroDyne Medical) to the participant's non-dominant index and ring fingers. Measurements were collected via the open-source e-Health biofeedback shield for biomedical applications (Libelium Comunicaciones) at a sampling rate of 100 Hz. Physiological data and audio markers were stored in plain text format and uploaded into a Matlab-based software platform for the analysis of skin conductance data (Ledalab). The analytical process is detailed in the results section below.

Stimuli were recorded at the beginning of each experimental session via Matlab. Participants were asked to sing and record a sustained pitch (approximately 3 seconds long) near the middle portion of their vocal register. After recording this tone, the spectral envelope of the participant's sung pitch was re-synthesized using the Matlab-based TANDEM-STRAIGHT vocoder. Four unique spectral envelope ratio (SER) manipulations were applied to the recorded tone; these manipulations were nominally labeled as very small (SER x 0.7143), small (SER x 0.833), large (SER x 1.2) , and very large (SER x 1.4) stimuli. Visualizations of the stimuli can be found in figure 1.

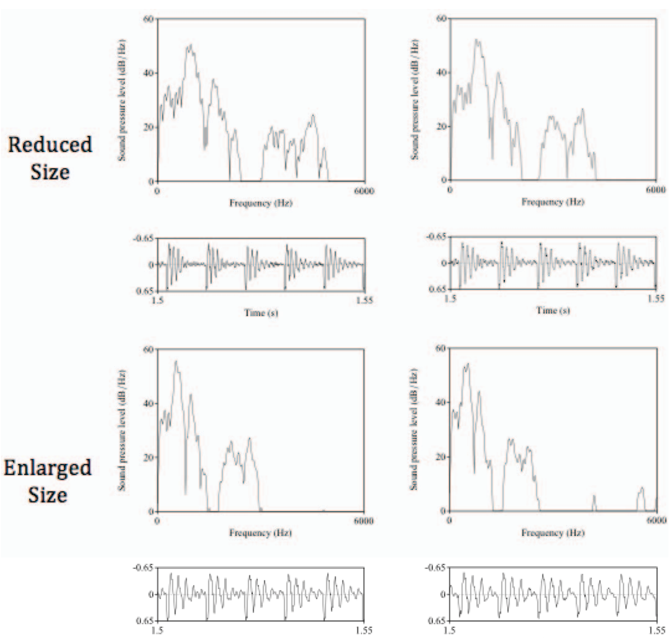


Figure 1a-d. Example waveforms and spectral-slices for each of the four SER manipulations used in the study. Spectral modifications were accomplished via the TANDEM-STRAIGHT vocoder. Figure 1a (upper-left) = very small, Figure 1b (upper right) = small, Figure 1c (bottom-left) = large, Figure 1d= very large.

In order to quickly generate experimental trials for each participant, we used the SoX command line audio tool to concatenate their resynthesized tones. Stimulus order within each trial was quasi-randomized via 24 unique concatenation scripts; these scripts ensured that any given stimulus was not presented consecutively, and that all four stimuli appeared in each ordinal position. Following a paradigm similar to that of Olsen & Stevens (2013), each of our trials consisted of 8 stimuli (i.e., each stimuli was presented twice) with approximately 15 seconds of silence between each stimulus.

C. Procedure

We collected baseline skin conductance data for a period of two minutes before presenting each participant with their uniquely generated trial. Participants were seated in a comfortable chair and instructed to refrain from moving their non-dominant hand. We then read them the following instructions:

"You will now hear a collection of sounds. While you are listening, all you must do is relax and keep your non-dominant hand still. This portion of the experiment will last approximately 4:30."

After listening to the complete trial, we removed the skin conductance electrodes, and then participants once again listened to their uniquely generated trial. This time, participants provided spoken feedback to indicate if the stimuli were perceived as originating from a larger or smaller sound source, relative to themselves. The experimenter transcribed these data points directly into a spreadsheet.

III. Results

When assessing the perceived size of each sound, on average participants correctly identified the sound source size as being relatively larger or smaller than themselves 85.05% (SD=15.99%) of the time. Of the 29 participants, 11 correctly assessed the size of all 8 sounds, 7 misidentified only 1 sound, 4 misidentified 2 sounds, 6 misidentified 3 sounds, and 1 participant misidentified 4 sounds. No participants misidentified 5 or more sounds.

Skin conductance recordings were analyzed using Ledalab. The majority of our participants were unfortunately deemed to be non-responders. That is, their phasic skin conductance data did not reveal any significant SCRs. Our primary conjecture for this finding was that the room in which our experiment was conducted was relatively cold and potentially inhibited participants' normal electrodermal activity. Methodological studies on skin conductance recording suggest that measurements are optimal with an ambient temperature between 22-24℃ (Braithwaite et al., 2015; MacNeill & Bradley, 2016); our lab space fluctuates between 17-19℃. Therefore, we reviewed our initial dataset and removed non-responders via visual inspection of the data. Because we are here reporting on a pilot study, we will offer a few modest descriptive findings from our small subset of responders (n = 5).

Each participant's trial data were processed with Ledalab's adaptive smoothing feature followed by a Continuous Decomposition Analysis (Benedek & Kaernbach, 2010b). Prior to analysis, we ensured that audio markers for the onset of each stimulus were correct. The Continuous Decomposition Analysis (CDA) then analyzed a response window from 1s to 4s after each stimulus onset. The relevant event data that we chose to incorporate into our statistical models included the sum of significant skin conductance response (SCR) amplitudes, the maximum value of phasic activity, and the response latency of the first significant SCR within the response window. Due to the small dataset, we reduced the four factors (much smaller, smaller, larger, much larger) of our independent variable (relative size manipulation) into two groups (smaller & larger).

Our three skin conductance variables of interest are plotted as a function of size condition (i.e., smaller & larger) in figures 2a-c. From our pilot data, two of the variables were skewed in the direction predicted by our hypothesis. Specifically, both the sum of SCR amplitudes data and the phasic maximum data tended to increase as SER manipulations of size increased, as would be predicted by our hypothesis. However, it bears reminding that these results are based on a very small sample of responders. We also expected that SCR latency times would be shortest for sounds that were perceived to be larger. However, these data were skewed opposite of our predicted direction.

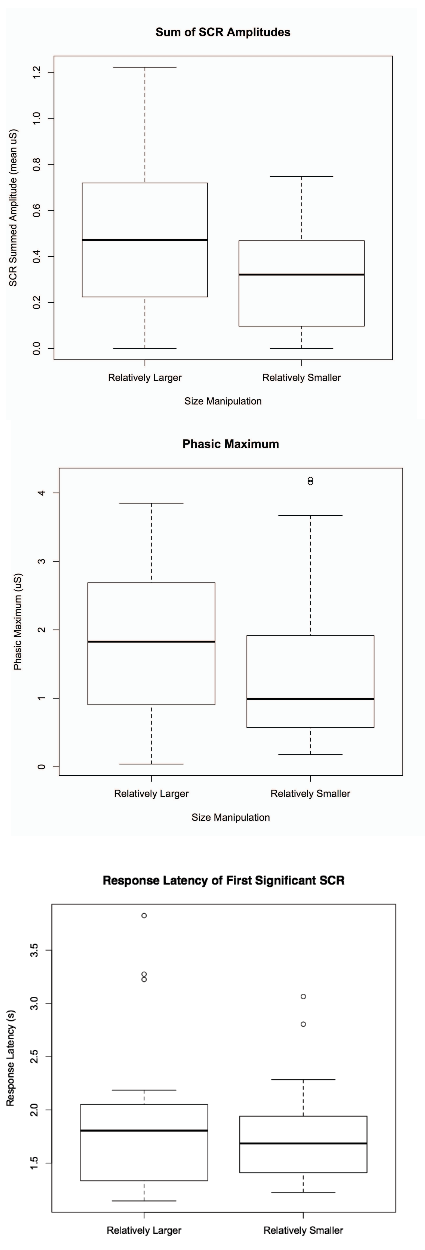


Figure 2a-c. Three SCR measurements as a function of experimental stimuli (relatively smaller or larger than each participant).

IV. DISCUSSION

Our goal for this pilot study was to determine if listeners would be capable of classifying sounds as being larger or smaller than themselves, and further, to determine if relatively larger sounds would be more likely to result in induced emotional responses. The results, although preliminary, warrant further research. On average, participants in our study were 85% accurate when judging sound source size relative to their own size. For our small subset of responder participants, we observed a positive relationship between relatively larger sound sources and both summed SCR amplitudes and SCR phasic maximums. We did not observe shorter SCR latency times for larger stimuli.

We believe that our results from this pilot study provide a novel insight towards understanding induced musical affect by "normalizing" sound stimuli to each specific listener. This approach, combined with a better understanding of how sound source size is inferred for instrumental timbres, has a number of musical implications. First, listener-normalized emotional responses might reveal how various orchestrations, from sparse to grandiose, can reliably induce certain types of emotional responses via music. Secondly, listener normalized emotional responses might be useful for understanding certain types of affective human computer interactions. Finally, listener-centered approaches to emotional communication within music might inform us about variability in previously conducted studies of induced emotion.

ACKNOWLEDGMENTS

The authors both wish to thank Illinois Wesleyan University for providing financial support to realize this research. Joe Plazak wishes to thank both Stephen McAdams and Marc Pell (McGill University) for their support throughout his research leave.

REFERENCES

Benedek, M., & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of neuroscience methods, 190*(1), 80-91.

Boucsein, W. (2012). *Electrodermal activity (2nd ed.).* New York: Springer Science+Business Media, LLC.

Braithwaite, J.J., & Watson, D.G (2015). Issues surrounding the normalization and standardisation of skin conductance responses (SCRs). Technical Research Note. Selective Attention & Awareness Laboratory (SAAL), Behavioural Brain Sciences Centre, School of Psychology, University of Birmingham.

Clarke, E. F. (2005). *Ways of listening: An ecological approach to the perception of musical meaning*. Oxford University Press.

Gabrielsson, A. (2002). Emotion perceived and emotion felt: Same or different?. *Musicae Scientiae, 5*(1 suppl), 123-147.

Gaver, W. W. (1993). What in the world do we hear?: An ecological approach to auditory event perception. *Ecological psychology*, *5*(1), 1-29.

Juslin, P. N., & Laukka, P. (2004). Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of New Music Research, 33*(3), 217-238.

Kawahara, H., Takahashi, T., Morise, M., & Banno, H. (2009). Development of exploratory research tools based on TANDEM-STRAIGHT. In *Proceedings: APSIPA ASC 2009: Asia-Pacific*

Signal and Information Processing Association, 2009 Annual Summit and Conference, 111-120.

MacNeill, A. L. and Bradley, M. T. (2016), Temperature effects on polygraph detection of concealed information. *Psychophysiology, 53*(2), 143-150.

Olsen, K. N., & Stevens, C. J. (2013). Psychophysiological response to acoustic intensity change in a musical chord. *Journal of Psychophysiology 27*(1), 16-26.

Plazak, J. (in press). Listener-normalized musical affect: conceptualizing "felt" musical emotions.

Rickard, N. S. (2004). Intense emotional responses to music: a test of the physiological arousal hypothesis. *Psychology of Music*, *32*(4), 371-388.

Schubert, E. (2013). Emotion felt by the listener and expressed by the music: literature review and theoretical perspectives. *Frontiers in Psychology* 4:837.

The Effect of Repetitive Structure on Enjoyment in Uplifting Trance Music

Kat Agres,[*1] Louis Bigo,[#†2] Dorien Herremans,[*3] Darrell Conklin[#§4]

[*]*Center for Digital Music, Queen-Mary University of London, United Kingdom*

[#]*University of the Basque Country, UPV/EHU, Bilbao, Spain*

[†]*Institute for music/acoustic research and coordination (IRCAM), Paris, France*

[§]*IKERBASQUE, Basque Foundation for Science, Bilbao, Spain*

[1]kathleen.agres@qmul.ac.uk, [2]louis.bigo@ircam.fr, [3]d.herremans@qmul.ac.uk, [4]darrell.conklin@ehu.es

ABSTRACT

Whereas previous studies have examined the effect of rhythmic structures in trance music, the present research explores the impact of harmonic repetition on enjoyment through empirical testing. A number of uplifting trance (UT) excerpts were generated with different semiotic patterns (structures defining the order of chords in the sequence) each with varying levels of harmonic repetition. Listeners then provided enjoyment ratings for each of the excerpts. The results of the experiment indicate that harmonic repetition, not only rhythmic or percussive structure, does in fact contribute to the enjoyment of uplifting trance music. Depending on the harmonic context, greater chord diversity or maximal repetition elicits high enjoyment.

I. INTRODUCTION

Across cultures, various forms of trance music are often used to evoke alternative listening states, such as heightened enjoyment and a sense of becoming "lost in the music" (Sacks, 2006). Repetition is thought to be crucial for evoking these listening states (Walsh, 1989), yet the particular repeated elements influencing affective response in listeners remain unclear. Previous research focuses upon the intuitive connection between rhythmic/percussive elements and the physiological entrainment underpinning heightened enjoyment (Becker-Blease, 2004; Becker, 2012; Fachner, 2011; Neher, 1962; Trost, Frühholz, Schön, Labbé, Pichon, Grandjean & Vuilleumier, 2014), but we focus on a less understood element of trance music: harmonic repetition. We aim to elucidate how local harmonic structure (e.g., chord progressions) contributes to the experience and enjoyment of 'uplifting trance' (UT) music, a sub-genre of electronic dance music characterized by repetitive melodies, chord sequences, and rhythmic patterns (Madrid, 2008). UT pieces have a particular musical structure that includes distinctive functional sections known as the *breakdown*, the *buildup*, and the *anthem*. This study places the most distinctive and fundamental elements of a UT piece in focus: the anthem, arguably the energetic height of the UT listening experience. The next section describes a behavioral experiment that was conducted in order to test the relationship between local harmonic structure and listeners' enjoyment.

II. BEHAVIORAL EXPERIMENT

A. Hypotheses

To elucidate the connection between repetition of harmonic structure and subjective enjoyment, we conducted a behavioural experiment in which listeners provided enjoyment ratings for UT anthems varying in repetitiveness of chord sequences. Two alternative hypotheses were considered: 1) Listeners display a preference for repetitive harmonic structures (more complex patterns yield less enjoyment, as shown in the left plot of Figure 1), and 2) Enjoyment reflects the inverse-U relationship between preference and complexity predicted by the Wundt curve (Berlyne, 1970) (shown in the right plot of Figure 1), which has been demonstrated for other musical genres (Heyduk, 1975; Huron, 2006; North & Hargreaves, 1995). Note, however, that very complex stimuli (those that should yield very low enjoyment ratings according to the Wundt curve) were not included in this study, because the research aimed to maintain as much ecological validity as possible, and because generating very complex sequences was often not possible given the simple semiotic structures employed (see next section for details). As a result, this research addresses a portion of the Wundt hypothesis (outlined in blue in Figure 1): that greater chord diversity in the stimuli (higher complexity / lower repetitiveness) generally results in greater enjoyment.

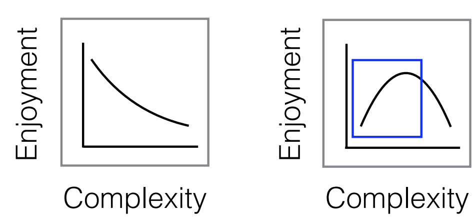


Figure 1. Depicted on the left is Hypothesis 1, that enjoyment decreases as stimuli become less repetitive and more complex. Depicted on the right is Hypothesis 2, which displays the relationship between preference and complexity as predicted by the Wundt curve. Because no very complex stimuli were included in this research, the study only addresses the portion of the curve outlined by the blue box.

B. Method

In the behavioral experiment, listeners provided enjoyment ratings for UT excerpts varying in harmonic repetition. The sequence of chords within each stimulus was determined by an underlying semiotic pattern, and a total of 14 different semiotic patterns were used to constrain harmonic repetition (see Table 1). The symbols A, B, etc, do not refer to explicit chord names, but are labels indicating the order and repetition of chords within the semiotic patterns. Each label corresponds to a chord that has a duration of two bars, and therefore every pattern has a duration of 16 bars.

Table 1. The 14 semiotic structures used in the experiment.

AABB-AAAA	ABCD-ABCD
AABB-AABB	ABCD-AECF
AABB-AACC	ABCD-BCDB
AABB-BBAA	ABCD-CECF
AABB-CCAA	ABCD-DADB
AABB-CCAB	ABCD-EFAB
AABB-CCDE	ABCD-EFGH

The semiotic patterns were created such that the first 8 bars were either ABCD or AABB. The second 8 bars fell into three categories when considering the particular chord *and* its metrical position in the pattern: either the semiotic structure of the first half was completely preserved (as in AABB-AABB and ABCD-ABCD), some of the semiotic structure of the first half was preserved (e.g., **AA**BB-**AA**CC and AB**C**D-CE**C**F, where the repeated chords are in bold), or none of the semiotic structure of the first half was preserved (e.g., AABB-CCDE and ABCD-DADB). Finally, the number of unique chords in the second half of the semiotic structure ranged from 1 chord (AAAA) to 4 chords (ABCD or EFGH). The semiotic structures used in the listening experiment are displayed in Table 1.

The listeners were presented with a total of 56 UT anthems (4 instances per unique semiotic pattern), each 30 seconds in duration, and the order of trials was randomised across participants. Two practice trials were played before the experiment to help the participant become familiar with the experimental procedure. After listening to each trial, the listeners' task was to provide enjoyment ratings on a 7-point Likert scale, where 1 was equivalent to 'Did not enjoy at all' and 7 represented 'Enjoyed very much.'

The experiment was conducted on a MacBook Pro laptop using a GUI constructed for data collection for this experiment. Participants listened to the stimuli through headphones set at a comfortable listening volume, in a quiet room.

C. Participants

Twenty volunteers, recruited via email advertisement and flyers around the Queen Mary University of London campus, participated in the experiment (mean age = 28.6 yrs, std = 7.9 yrs; 6 female and 14 male). Participants included undergraduate students, graduate students, and staff, all of whom reported prior experience listening to trance music. Each volunteer received £7 for his or her participation.

D. Stimulus generation

To avoid any potential confound of the actual chords presented (as opposed to the overarching semiotic structure), four different stimuli were generated for each semiotic structure listed in Table 1. To this end, the first chord of each sequence was either Bmin, Dmaj, Emin or Gmaj. The generation was performed using a method which allows sampling high probability chord sequences with respect to a given semiotic structure (Conklin, 2015). Sequences were generated from a statistical model that encodes chord transition probabilities of a corpus of 100 uplifting trance chord loops designed specifically for this study [1]. The sequences with the highest probability between successive chords were selected and key modulations within the sequences were avoided as much as possible. Each chord sequence was then rendered within the Digital Audio Workstation Logic Pro X (LPX) by starting from an existing uplifting trance template[2] and applying as few as possible pitch modifications necessary to make its harmonic structure fit with the generated chord sequence (Bigo & Conklin, 2015). As this transformation task leaves rhythm, instrumentation, and audio effects unchanged, the generation results in a set of stimuli that differ by their harmonic properties. The entire generation process was automated by using a system which interacts with LPX through the MIDI protocol (Conklin & Bigo, 2015).

E. Results

To test the impact of repetition on enjoyment ratings, a 2 X 3 X 4 ANOVA was conducted for three types of repetition found within the semiotic structures as described above. These three repetition variables were: 1) the structure of first 8 bars (AABB or ABCD), 2) the number of chords in common, and in the same metrical position, between the first and second 8 bars of the sequence (all the same, partially the same, or no common chords), and 3) the number of unique chords in the second 8 bars (ranging from 1 to 4 chords).

There was a significant effect of variable 3 (the number of unique chords in the second 8 bars), $F = 4.24$, $p < 0.01$, such that more chord diversity yields higher enjoyment ratings. A significant interaction was also found between variable 1 (semiotic structure of first 8 bars) and variable 2 (all, some, or no chords in common between the first and second 8 bars), $F = 4.09$, $p < 0.05$. Highest enjoyment ratings were elicited when all the same chords were present in the first and second half of the stimulus (AABB-AABB and ABCD-ABCD). For stimuli beginning with ABCD, high enjoyment ratings were also found when none of the same chords were present in the first and second half of the stimuli. The lowest enjoyment ratings were elicited when the second 8 bars contained only some of the same chords as the first 8 bars, especially for stimuli beginning AABB. Lastly, a post hoc analysis indicated that semiotic patterns occurring in the Uplifting Trance Anthem Loops Corpus (i.e. native structures) were preferred over those not in corpus, $t = 2.36$, $p < 0.01$.

F. Discussion

Although there was a trend of increasing enjoyment for stimuli with less repetition in the last 8 bars (that is, more unique chords in the second half of the stimulus), this was not true for sequences beginning AABB. The findings indicate that enjoyment is elicited by semiotic patterns that are either very repetitive, namely, AABB-AABB and ABCD-ABCD, or fairly complex, such as ABCD-EFAB and ABCD-BCDB; but enjoyment does not result from sequences that are only slightly repetitive, such as AABB-AACC. Violations of

[1] The Uplifting Trance Anthem Loops Corpus (UTALC) which was constructed for this study may be found here: http://www.lacl.fr/~lbigo/utalc

[2] Uplifting Trance Logic Pro X Template by Insight, DAW Templates, Germany.

stylistic form (e.g., culminating a semiotic pattern with AA when a chord change is expected) were also disliked.

Overall, the results did not fully support or reject either of the two hypotheses; rather, we speculate that listeners' enjoyment reflects a combination of the two. Although evidence was not found across all semiotic patterns to support the first hypothesis (that greater repetition yields greater enjoyment), most of the highest rated sequences were the most repetitive sequences in the study (AABB-AABB and ABCD-ABCD). The second hypothesis, that enjoyment increases as complexity increases, up to a certain point (the climax of the Wundt curve, and the portion of the curve tested here), was also supported in part: stimuli containing only slightly repetitive chords were disliked compared with stimuli containing little to no repetition (that is, more complex harmonic structures produced higher enjoyment ratings than moderately complex structures). A schematic of this theoretical combination of hypotheses is depicted in Figure 2. These preliminary findings may suggest that listeners' preferences are divided, with greatest enjoyment elicited by either the most repetitive harmonic structures (the lefthand maximum of the pink box in Figure 2) or fairly complex, varied structures (the righthand maximum of the pink box). Again, no highly complex sequences were included, as this study focused only upon low to moderately complex stimuli.

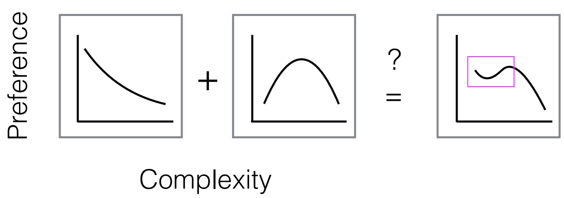


Figure 2. The experimental results are accounted for by a combination of the first two hypotheses.

III. CONCLUSION

By systematically investigating the repetitiveness of semiotic structures, we have discovered specific contexts in which chord repetitions influence the enjoyment of UT music. These preliminary findings suggest that harmonic repetition, not simply repetitive rhythmic and percussive structures, do in fact contribute to enjoyment of this genre. The authors are currently working on a follow-up study to further substantiate these results. Future work will explore the connection between repetitive harmonic elements and altered listening states signifying heightened enjoyment, such as audience flow.

ACKNOWLEDGMENT

This research is supported by the project Lrn2Cre8 which is funded by the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET grant number 610859. This project also received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement number 658914.

REFERENCES

Becker-Blease, K. A. (2004). Dissociative states through new age and electronic trance music. *Journal of trauma & dissociation*, *5*(2), 89-100.

Becker, J. (2012). 3 Rhythmic Entrainment and Evolution. *Music, Science, and the Rhythmic Brain: Cultural and Clinical Implications*, *1*, 49.

Berlyne, D. E. (1970). Novelty, complexity, and hedonic value. *Perception & Psychophysics*, *8*(5), 279-286.

Bigo, L., & Conklin, D. (2015). A viewpoint approach to symbolic music transformation. In *11th International Symposium on Computer Music Multidisciplinary Research (CMMR 2015)* (pp. 56-70).

Conklin, D. (2015). Chord sequence generation with semiotic patterns. *MML 2015: 8th International Workshop on Machine Learning and Music*, 1-3.

Conklin, D., & Bigo, L. (2015). Trance generation by transformation. *MML 2015: 8th International Workshop on Machine Learning and Music*, 4-6.

Fachner, J. C. (2011). Time is the key—Music and altered states of consciousness. *Altering consciousness: A multidisciplinary perspective*, *1*, 355-376.

Heyduk, R. G. (1975). Rated preference for musical compositions as it relates to complexity and exposure frequency. *Perception & Psychophysics*, *17*(1), 84-90.

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. Cambridge, MA: MIT Press.

Madrid, A. L. (2008). *Nor-tec Rifa!: Electronic Dance Music from Tijuana to the World: Electronic Dance Music from Tijuana to the World*. Oxford University Press, USA.

Neher, A. (1962). A physiological explanation of unusual behavior in ceremonies involving drums. *Human biology*, *34*(2), 151-160.

North, A. C., & Hargreaves, D. J. (1995). Subjective complexity, familiarity, and liking for popular music. *Psychomusicology: A Journal of Research in Music Cognition*, *14*(1-2), 77.

Sacks, O. (2006). The power of music. *Brain*, *129*(10), 2528-2532.

Trost, W., Frühholz, S., Schön, D., Labbé, C., Pichon, S., Grandjean, D., & Vuilleumier, P. (2014). Getting the beat: entrainment of brain activity by musical rhythm and pleasantness. *NeuroImage*, *103*, 55-64.

Walsh, R. (1989). Shamanism and early human technology: The technology of transcendence. *ReVision*, *21*, 34-40.

Emotions and chills at the interface between contrasting musical passages

Sabrina Sattmann[1] and Richard Parncutt[1]
[1]Center for Systematic Musicology, University of Graz, Graz, Austria

Which structural features in music evoke chills, and how are chills related to emotion? 20 participants each brought 3 chill-eliciting pieces of their own choice. In an interview, they indicated the chill-inducing passages verbally and evaluated the intensity of felt emotions using the Geneva Emotional Music Scale (GEMS-25; Zentner et al., 2008), complemented by items which were found to be important in a previous study. On average, each participant indicated 5.9 chill-inducing sections, making a total of 118 passages. On factor level, the emotion most strongly associated with chills was wonder, followed by power, transcendence and joyful activation, all of which were rated higher than tension and sadness. On item level, we found highest ratings for fascinated, strong, liberated (item added by authors), happy and allured, whereas emotions like sad, tense, soothed, nostalgic or tearful received very low ratings. This contradicts Panksepp (1995) who suggested that musical chills are associated with the emotions sad/melancholy and thoughtful/nostalgic. Analysis of the 118 passages showed that chills often occur after a relatively monotonous section, characterized by a repetition of harmony, rhythm and melody, a delay of progression, a reduced number of instruments or short silence. The chill-inducing passage itself featured a sudden increase in loudness, an entry or change of instrument(s), a melodic peak, a change in melodic range or a sudden rhythmical/harmonical change. In sum, the emotions associated with chills are attributed to two contrasting factors: wonder and transcendence (sublimity factor) on the one hand, power and joyful activation (vitality factor) on the other hand; this could give further insight into understanding the phenomenon of chills. As observed by Grewe et al. (2007), chills were often evoked by sudden changes; in addition, we found that the preceding passage was characterized by a noticeable lack of change or activity, suggesting that the chill was evoked by the contrast between the two passages.

The influence of odour on the perception of musical emotions

Scott Beveridge[1]
[1]Department of Engineering, Glasgow Caledonian University, UK

Affective evaluations of music can often involve integration of multiple sensory cues. Previous work in this area has focused on the role of visual cues that performers combine with acoustic cues to express particular emotions. These visual cues include facial expression and body movement/posture. This paper examines the role of olfactory stimuli in music emotion perception. Despite evidence that odours can effect general perceptual and cognitive processing, little work has been done in the context of music. Of particular interest is the possibility of affect-congruent judgment biases in relation to music. As valence is a primary dimension in characterizing odour, we are interested in the effect this may have on the perceived valence of music.

The aim of this study is to investigate the effect of odour on the perception of musical emotions. Specifically, we are interested in changes in strength of perceived emotions and the time taken to make emotion judgements.

30 participants were randomly assigned to either a control or experimental condition. In the experimental condition, a vanilla odour was diffused into the experimental environment. Vanilla was chosen as a stimulus as it is considered a positively valenced, pleasant odour. Participants were asked to rate their perceived emotions of 30 pieces film music. A software tool was used to rate each piece in terms of arousal and valence. The software also recorded the time taken for participants to make responses.

Preliminary findings show that in the odour condition, the strength of ratings for negatively valenced music is reduced. Furthermore, the response time of emotion judgements are significantly smaller in the odour condition.

This early study suggests that environmental odour can modulate emotional judgements of music. More specifically, sad music is perceived less sad in the presence of a positively valenced, pleasing odour. In addition, judgements of perceived emotions are made in less time in the presence of a pleasing odour.

Emotional mediation in crossmodal correspondences from music to visual texture

Thomas Langlois[1], Joshua Peterson[1], Stephen E. Palmer[1]
[1]Department of Psychology, University of California, Berkeley, CA, USA

Background
Previous work has revealed systematic music-to-color cross-modal associations in non-synesthetes that are mediated by emotion (e.g., Palmer et al., 2013; Langlois et al., VSS-2013; Whiteford et al., VSS-2013).

Aims
What identifiable features might mediate cross-modal associations from music to visual texture by using single-line melodies that vary along highly controlled musical dimensions: register (low/high), mode (major/minor), note-rate (slow/medium/fast), and timbre (piano-sound/cello-sound)? Are these associations mediated by emotional and/or non-emotional features?

Methods
Non-synesthetic participants made visual texture associations to each of 32 variations on 4 melodies by Mozart from a 7×4 array of black-and-white visual line-based textures.

Results
Cross-modal melody-to-texture associations were at least partly mediated by emotion, because the correlation between the emotional ratings of the music and the emotional ratings of the chosen textures were high (Angry/Not-Angry: r=.79, Calm/Agitated: r=.91, Active/Passive: r=.76). However, unlike music-to-color associations, the Happy/Sad correlation (r=.31) was not significant. Melody-to-texture associations were also mediated by sensory-perceptual features (e.g., Sharp/Smooth: r=.96, Curved/Straight: r=.92, Simple/Complex: r=.89, Granular/Fibrous: r=.80). Importantly, the cross-modal melody-to-texture associations corresponded to specific musical features. k-means clustering analyses revealed strong timbre and note-rate effects: cello melodies were paired with straight/sharp/fibrous textures, and piano melodies with curved/smooth/granular textures. Different textures were also chosen for different note-rates, with more granular/separate textures being chosen with slow note-rate melodies, and more fibrous/connected textures being chosen with faster note-rate melodies.

Conclusions
Both emotional and sensory-perceptual features are important for understanding melody-to-texture associations.

Expressive performance and listeners' decoding of performed emotions: A multi-lab replication and extension

Renee Schapiro[1], Jessica Akkermans[1], Veronika Busch[2], Kai Lothwesen[2], Timo Fischinger[3], Klaus Frieler[4], Kathrin Schlemmer[5], Daniel Shanahan[6], Kelly Jakubowski[1], and Daniel Müllensiefen[1]

[1]Psychology Department, Goldsmiths University of London, UK
[2]Institute for Musicology and Music Education, University of Bremen, Germany
[3]Music Department, Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, Germany
[4]University of Music "Franz Liszt" Weimar, Germany
[5]Musicology Department, Catholic University Eichstätt-Ingolstadt, Germany
[6]Music Theory, Louisiana State University, Baton Rouge, Louisiana, USA

Until recently, there has been a lack of replication studies conducted in the field of music psychology. A highly-cited paper in the field by Juslin and Gabrielsson (1996) reported that performer's intended emotional expressions were decoded by listeners with a high degree of accuracy. While there have been related studies published on this topic, there has yet to be a direct replication of this paper. The present experiment joins the recent replication effort by producing a multi-lab replication using the original methodology of Juslin and Gabrielsson. Expressive performances of various emotions (e.g., happy, sad, angry, etc.) by professional musicians were recorded using the same melodies from the original study and are subsequently being presented to participants for emotional decoding (i.e., participants will rate the emotional quality of each excerpt using a 0-10 scale). The same instruments from the original study have been used (i.e., violin, voice, and flute), with the addition of piano. Furthermore, this experiment investigates potential factors (e.g., musicality, emotional intelligence, emotional contagion) that might explain individual differences in the decoding process. Finally, acoustic features in the recordings will be analyzed post-hoc to assess which musical features contribute to the effective communication of emotions. The results of the acoustic and individual differences analyses will contribute to a more comprehensive understanding of when music can be an effective vehicle for emotional expression.

Self-chosen music listening at home, emotion and mood regulation

Alexandra Lamont[1] & Laura Vassell[1]
[1]School of Psychology, Keele University, Keele, Staffordshire, UK

In Western industrialised society, music listening occurs in a wide range of settings and serves a wide range of functions. Studies have begun to focus on specific contexts to explore how these might influence the nature of music listening (e.g. driving: Dibben & Williamson, 2007; travelling: Heye & Lamont, 2010). This presentation focuses on self-chosen music listening in the home as a specific context with potential for emotion regulation (Saarikallio, 2011): listeners have access to their entire music collection and are able to express and work through emotions more readily in private. The study explores the fine-grained emotions evoked by music listening using the novel technique of Subjective Evidence-Based Ethnography (SEBE; Glaveanu & Lahlou, 2012). This technique obtains first person digital-audio from a camera worn at eye level, which forms the starting point for an interview which allows participants to reflect on moment-to-moment experiences in greater depth than a typical retrospective interview can. Twenty young adults will wear the glasses and make recordings at home while they are listening to music over one week. Given the prevalence of music listening in everyday life, this will generate a large number of musical episodes, some longer and others shorter, that can be used as a prompt to engage participants in a detailed discussion of the emotions they experience through music and to situate this in the context of their everyday lives. Given the fleeting nature of many emotional responses to music and emotions experienced in music through the processes of self-regulation, this data will shed considerable light on the ways in which people use music in everyday life to regulate their emotions. Pilot testing is under way with the glasses to establish the precise nature of instructions to participants and early results indicate that music listening at home tends to comprise longer episodes than are typical in other settings.

Hearing Wagner: Physiological Responses to Richard Wagner's *Der Ring des Nibelungen*

David John Baker[1], Daniel Müllensiefen[2]
[1]Music Department, Louisiana State University, Baton Rouge, Louisiana, USA
[2]Psychology Department, Goldsmiths, University of London, London, UK

The relationship between music and emotion has frequently been studied either in controlled settings or through retrospective reports. However, these approaches lack the long-term perspective of emotional developments across real concert experiences and often require participants to disengage from the concert experience. In contrast, this paper presents results from two experiments in which participants listened to the entirety of Richard Wagner's *Der Ring des Nibelungen*, while being continuously assessed for their arousal levels. We intend to explore how unconscious arousal affects perception and memory for musical materials, namely leitmotives in *The Ring*, as well as use the physiological data to explore musicologically motivated questions. The first experiment was conducted live in a large opera house (N=7) and the second group screened condition (N=9). Physiological reactions from all participants were measured via their heart rate, electrodermal activity, and small scale movements. In addition, individual differences measures such as musical sophistication, Wagner affinity and knowledge, auditory working memory span and personality were taken from participants. Participant's familiarity and recognition memory for individual leitmotives was taken prior to each opera. Results from the live experiment suggest that memory for musical material increases with exposure time for both novices and experts of Wagner's music alike, although participants with a higher Wagner affinity generally show stronger arousal reactions. The strongest reactions from both groups were observed for the opera *Siegfried*. Further results, including comparisons between participants in live and screen conditions and results regarding individual leitmotives are being analysed. The results suggest that given the differences in physiological and behavioral responses, factors beyond musical training should be further investigated to better understand the listening experience.

The effects of hearing aids and hearing impairment on the physiological response to emotional speech

Gabriel A. Nespoli[1], Gurjit Singh[1,2], and Frank A. Russo[1]
[1]Ryerson University, Toronto, Ontario, Canada
[2]Phonak AG, Mississauga, Ontario, Canada

Age-related hearing loss negatively impacts the perception of speech. While hearing aids can ameliorate these deficits to some extent, questions exist about their efficacy for supporting perception of emotion. In normal hearing adults, characteristic physiological responses are triggered in response to speech emotion. The current study assessed the extent to which these responses vary across normal hearing, hearing impaired, and hearing aided older adults. Participants were presented with audio-only samples of linguistically neutral sentences spoken in a happy, sad, angry, or calm manner, and were asked to respond with the expressed emotion. Physiological responses included skin conductance, heart rate, respiration rate, and facial electromyography on cheek (smiling) and brow (frowning) muscles. Normal hearing participants were both faster and more accurate in their responses than the hearing-impaired or hearing-aided groups. Normal hearing participants exhibited an increase in skin conductance in response to negatively-valenced and high-arousal stimuli (i.e. angry, and to some extent happy and sad). This increase was not present in hearing-impaired participants but was recovered in hearing-aided participants. These findings raise important questions about the efficacy of signal processing strategies employed in modern digital hearing aids.

Coping with interpersonal distress: Can listening to music alleviate loneliness?

Katharina Schäfer[1] and Tuomas Eerola[1,2]
[1] Department of Music, University of Jyväskylä, Jyväskylä, Finland
[2] Department of Music, Durham University, Durham, UK

Loneliness has been identified as a major risk factor for our health, but it can be influenced positively by the use of media. In order to feel connected to others, people often immerse emotionally in narratives or form para-social relationships with media figures (Hawkley & Cacioppo, 2010). There is some empirical evidence (Derrick et al., 2009; Saarikallio & Erkkilä, 2007) that music could act as a “social surrogate”.

To test the effect of music listening on loneliness, a between-subject design with two factors was implemented: (i) The need to belong, which comprised three conditions of autobiographical recall (interpersonal distress, task-related distress, and control), and (ii) an imagined music listening situation, which included either preferred or casual music. Both factors were operationalized as writing tasks. Mood was assessed before, between and after the writing. After the tasks, measures of loneliness and attachment were collected.

The results of 141 participants indicate that preferred music does not buffer against the negative emotional effects of interpersonal distress. On the contrary, emotional loneliness was heightened through the imagination of a situation where one would listen to one’s preferred music. Yet at the same time, those who were thinking of their preferred music were in a significantly better mood. So, even though preferred music boosted feelings of loneliness, it also raised the mood simultaneously.

These intriguing findings can be explained from the perspective of emotion regulation: Thinking about interpersonal distress may have guided the participants to choose music that promotes insights and clarification. This ‘mental work’ (Saarikallio et al., 2007) may have enhanced the feeling of loneliness, but helped them to deal with that conflict, which raised their mood. So, even if music enhanced loneliness in the short term, it could contribute to mental health in the long run as it supports healthy coping.

Does the Familiarity and Timing of the Music Affect Indirect Measures of Emotion?

L. Edelman, P. Helm, S.C. Levine, S.D. Levine, G. Morrello

Abstract— Extensive research has concluded that music can induce emotions in a listener. Much of the research has been based on self-report measures. Research suggests indirect measures avoid the biases inherent in self-report. The current study uses an indirect measure (word stem completion task) to look at the subconscious effect of music on emotions. We hypothesized that when participants hear happy, sad or scary music, they would be primed to generate words associated with the corresponding emotion. In addition, we added a familiarity condition including a familiar and unfamiliar song for each emotion. We hypothesized that participants would have stronger emotional responses to more familiar music. Results showed that the tone of the music did affect the amount of related emotion words written.

Keywords— Indirect measures, emotions, familiarity,

I. INTRODUCTION

There has been extensive research confirming that music can induce emotions in listeners [1]. Researchers have used a wide variety of response measures. These measures have included self-report rank ordering of emotional responses [2], standard emotion scales [3], scales developed specifically for music [4]. The self-report scales have been criticized for being confounded by demand characteristics [5]. Participants may guess what the researcher is looking for and respond accordingly. Others have used physiological responses, which are thought to be less influenced by demand characteristics [6]. Physiological measures are good for differentiating the intensity of the emotional response, however, they do not clearly differentiate the tone of the emotional response [7]. A previous researcher suggested the use of indirect measures of emotion to avoid demand characteristics and retain the emotional tone of the participants' responses [5].

Previous research has shown that a word stem completion task can be used as an implicit measure of emotional responses to happy and sad music [8]. However, across three studies, the word stem completion task only worked as a response measure when the music was familiar.

The current study expands the use of the word stem completion task by comparing music with three different emotional tones; happy, sad, scary, and varying the familiarity of the music. Additionally, since the music can be thought of as *priming* the participants to think about a particular emotion we varied whether the music was played before or during the word stem completion task. Our hypotheses were that participants would form word completions that matched the emotional tone of the music, and that the effect would be stronger for more familiar music. We reasoned that playing the music before participants started the task would have them primed before they started on the first words and therefore the results would be more clear cut than having them start the task before they heard much of the music.

II. METHOD

A. Participants

The sample included 134 participants all over the age of 18.

B. Design

A 3X2X3 mixed design was used. The between subjects independent variables were the emotional tone of the music; happy, sad, or scary, and the familiarity of the music (familiar or unfamiliar). The within subjects independent variable was the potential emotional tone of the word stems either happy, sad, or scary. The dependent measures were the completion of the word stems, and ratings of the music.

C. Stimuli

We used instrumental music to avoid associations with the lyrics:

- Happy familiar: *Build Me Up Buttercup* by Pentatonic Karaoke Band
- Happy unfamiliar: *Happy Whistling Ukulele* by Seastock
- Sad familiar: *Arms of an Angel* by The O'Neill Brothers Group
- Sad unfamiliar: *Hazel Emergency* from the original score of *The Fault in our Stars*
- Scary familiar: *Jaws* by The City of Prague Philharmonic
- Scary unfamiliar: *The Shadow* from the original score of *Psycho*

D. Materials

The materials consisted of a page of 20 word stems, which provided the first two letters of the word to be completed. All of the word stems had potential completions that were neutral in emotional tone. There were five words stems that had emotionally congruent completions for each of the three emotional tones of the music and a final five that had only neutral completions.

In the post-test questionnaire participants rated the music on its emotional tone, familiarity, liking, and how soothing, relaxing, and pleasant they found it. All of these were rated on a seven point scale with higher numbers meaning more. Participants also indicated whether they believed the music influenced their performance on the word stem task and whether or not they had musical training. If participants indicated that the music did influence their responses they were asked to explain how the music influenced them.

L. Edelman is with the Psychology Department, Muhlenberg College, Allentown, PA 18104 USA (phone: 484-664-3426; e-mail: ledelman@muhlenberg.edu).

P. Helm is with the Music Department, Muhlenberg College, Allentown, PA 18104 USA, (e-mail: pbhelm@muhlenberg.edu).

S.C. Levine, S.D. Levine, & G. Morrello are students in the Psychology Department Muhlenberg College, Allentown, PA 18104 USA.

E. Pretest

Previous researchers played the music while participants completed the word stems [8]. We hypothesized that priming the participants with the music before the word stem task would strengthen their responses. We ran the happy and sad conditions by playing the music either before or during the word stem task. There were no statistically significant differences in the results based on when the music was played. Therefore we decided to shorten the experiment by playing the music during the word stem task.

F. Procedure

Participants were told we were investigating the effects of music on word processing. They were given a word stem completion task while listening to either happy, sad, or scary music (either familiar or unfamiliar) and then given the post-test questionnaire to complete.

III. RESULTS

The data were analyzed using a 3 way mixed analysis of variance. There was a significant interaction between the tone of the music and the tone of word stem completion, $F(4,244)= 10.22$, $p<.01$, $\eta_p^2 = .14$. Fig.1 shows word stem completion was congruent with the tone of the music that was playing.

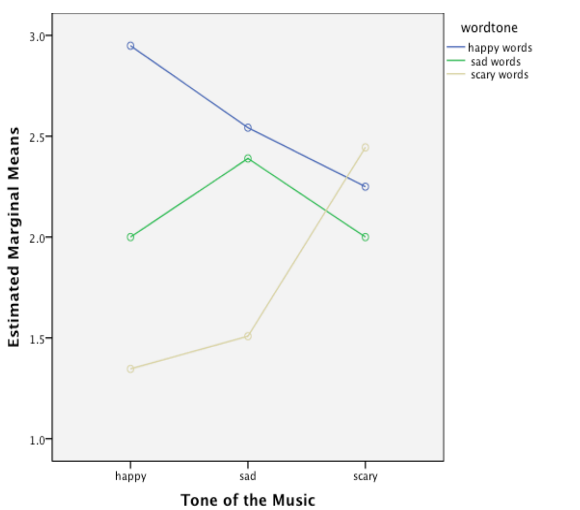


Fig. 1 Shows the average number of emotionally toned words completed by the emotional tone of the music.

We ran several manipulations checks. The pieces categorized as familiar were rated as more familiar (M= 5.68, SD=1.80) than the unfamiliar pieces, M=4.02, SD=1.96, F (1,167)= 39.23, $p<.01$, η_p^2 =.19. The emotional tone ratings significantly matched the categorization of the music, $F(2,167)= 216.05$, $p<.01$, $\eta_p^2= .72$.

We also ran correlations between the number of words completed for each emotional tone and the ratings of the music. The resulting correlations support the hypothesis that the perception of the music influenced participants' word stem completion. Participants that completed more happy words, rated the music as more familiar, ($r(173)= .26$, $p<.01$),more happy, ($r(173) = .16$, $p= .03$), more relaxing, ($r(173) = .26$, $p< .01$), more soothing, ($r(173) = .22$, $p< .01$), liked it more, ($r(173)= .22$, $p< .01$), and found it more pleasant $r(173)= .26$, $p<.01$. There were no significant correlations based on the sad words completed. Participants that completed more scary words, rated the music as less familiar, ($r(173)= -.17$, $p=.02$), more scary, ($r(173) = .24$, $p< .01$), less relaxing, ($r(173) = -.29$, $p< .01$), and found it less pleasant $r(173)= -.25$, $p<.01$.

There were several potential problems with our music choices. Although the pretest data indicated clear categorization for each piece of music, the ratings by the participants in the word stem completion task did not match the pretest ratings. There was a significant difference in liking based on the emotional tone of the music. Happy music was liked much more (M= 6.0, SD= 1.17) than the sad (M= 4.88, SD= 1.74) or scary music, M= 3.75, SD= 1.84, F (1, 167)= 27.06, $p<.01$, $\eta_p^2 = .25$. Happy music was rated as more familiar (M= 5.80, SD=1.42) than either the sad music (M= 4.02, SD= 2.04) or the scary music, M=3.75, SD= 2.25, $F(2,167)= 23.97$, $p<.01$, $\eta_p^2 = .22$. Although the unfamiliar music was always rated as less familiar, the difference in familiarity (between the unfamiliar and familiar pieces) varied by the tone of the music. Thus we were unable to draw any conclusions based on the role of familiarity.

IV. DISCUSSION

Because pretesting did not show a significant difference between whether participants heard the music before or during the word stem completion task, this condition was not included in a final analysis of the results. However, the results of the main study supported our hypothesis that indirect measures of emotion can be used in studies of musically induced emotions. We were able to confirm, via the tone of the word stems completed, that happy sounding music induced positive emotions within participants, sad sounding music induced negative emotions, and scary music induced scary emotions. The correlations also support this finding as they showed that the participants' ratings of the music played a role in the word stem completion task. We expected the emotional tone ratings to correlate with the word stem completions, but were surprised that that other ratings such as relaxing, soothing, and pleasant, influenced the word-stem completion task. Our results for familiarity were less clear. We had several confounds for our music categorizations. For example, the happy unfamiliar music rated higher on familiarity than the sad and scary music in both the familiar condition and unfamiliar conditions. The finding that the happy music was liked more than the sad or scary music may have also influenced the results.

Another limitation in our study may be in our sad music choices. For both the happy and the scary music there were correlations between the word stem completions and the participants' ratings of the music. However, there were no correlations between the number of sad words completed and the ratings of the sad music. This seems to indicate that our sad music choices did not work for our participants. Future researchers should use songs that are a better representation of each condition, especially in regards to familiarity.

IV. References

[1] Juslin, P. N., Liljeström, S., Västfjäll, D., & Lundqvist, L.-O. (2010). How does music evoke emotions? Exploring the underlying mechanisms. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications.*, Series in affective science (pp. 605–642). New York, NY US: Oxford University Press.

[2] Collier, G. L. (2007). Beyond valence and activity in the emotional connotations of music. *Psychology Of Music*, *35*(1), 10-31. doi:10.1177/0305735607068890

[3] Hunter, P.G., Schellenberg, E. G., & Schimmack, U. (2008). Mixed affective responses to music with conflicting cues. *Cognition And Emotion*, *22*(2), 327-352. doi:10.1080/02699930701438145

[4] Pearce, M. T., & Halpern, A. R. (2015). Age-related patterns in emotions evoked by music. *Psychology Of Aesthetics, Creativity, And The Arts*, *9*(3), 248-253. doi:10.1037/a0039279

[5] Västfjäll, Daniel (2011). Indirect perceptual, cognitive, and behavioural measures. In Patrik N. Juslin & John Sloboda (eds.), Handbook of Music and Emotion: Theory, Research, Applications. OUP Oxford.

[6] Lundqvist, L.-O., Carlsson, F., Hilmersson, P., Juslin, P. N. (2009). Emotional responses to music: Experience, expression, and physiology. *Psychology of Music*, *37*(1), 61–90.

[7] Gomez, P., & Danuser, B. (2007). Relationships between musical structure and psychophysiological measures of emotion. Emotion, 7, 377–387.

[8] Edelman, L., Baron, C., Rybnick, E., & Zimel, C. (2015) Music and Emotion: Indirect Measures. Poster presented at SMPC, Nashville, TN.

Emotional reactions in music listeners suffering from depression: differential activation of the underlying psychological mechanisms

Laura S. Sakka[1]
[1]Department of Psychology, Uppsala University, Uppsala, Sweden

Music has the potential to elicit emotions in listeners, and recent research shows that these emotions are mediated by a number of underlying psychological mechanisms, such as contagion and memories. Depression has been associated with negative biases in cognitive processes (e.g., memory, attention). The aim of this study was to explore whether such cognitive biases also influence the activation of mechanisms during music listening in depressed people. Depressed and healthy-control individuals participated in a music listening experiment, which featured musical stimuli that targeted specific mechanisms (i.e., brain stem reflex, contagion, and episodic memory). After each piece, listeners rated scales measuring the BRECVEMA mechanisms and their felt emotions. Participants also completed psychometric tests of depression (BDI) and anxiety (BAI). Depressed people showed differences in emotional reactions to music compared to healthy controls, and these could be partly explained by a differential activation of the mechanisms. These results suggest that cognitive biases may be evident in the activation of mechanisms during music listening, influencing the way depressed people experience music. Knowledge of this kind may be of potential use for music-therapeutic interventions against depression, where focus lies on the underlying mechanisms mediating the induction of musical emotions.

Music liking, arousal and cognition

Tessa Overboom[1], Jade Verheijen[2], Joyce de Bruin[3], Rebecca S. Schaefer[4]

Leiden Institute for Brain and Cognition / Health, Medical and Neuropsychology Unit, Leiden University, Leiden, The Netherlands

[1]tessamariaoverboom@gmail.com, [2]jade.v.jel@gmail.com, [3]joyce_de_bruin@hotmail.com, [4]r.s.schaefer@fsw.leidenuniv.nl

ABSTRACT

Findings suggesting that music listening can lead to improved cognitive performance have been interpreted as being driven by arousal, as a result of music preference. However, this idea has not been tested experimentally, and it is currently unclear how music preference, self-reported arousal (induced by and perceived in music), physiological arousal and familiarity affect short-term cognitive performance. The current study investigates the relationship between music preference, arousal (perceived, induced, and physiological, measured through heart rate) and cognitive performance using a mental rotation task. Preliminary analyses indicate that when contrasting different pieces of music, no relationship is found between differences in preference or arousal and differences in cognitive performance, however differences in liking are highly related to differences in familiarity and induced emotion, and differences in self-reported induced arousal are related to differences in heart rate during the cognitive task, supporting the current design and stimuli that were used. Although data collection still continues, the current analyses suggest that even if cognitive performance increases after music listening, which is not tested in the current experiment, the effects do not appear to be mediated by either music preference or arousal. If these relationships remain stable as data collection continues, these claims may be further specified. The findings may inform the design of interventions aimed at cognitive improvement, for instance in dementia care.

I. INTRODUCTION

In previous studies it has been suggested that music listening has a positive effect on cognitive performance. One of the most well known examples is the so-called Mozart effect (Rauscher, Shaw & Ky, 1993), which was subsequently interpreted to be related to music preference, and preference-induced arousal. Specifically, it is the musical stimulus that improves mood in the listener and perhaps arousal, leading to improved performance on various cognitive tests (cf. Schellenberg, 2014). However, this interpretation has not yet been experimentally evaluated.

Music has been found to lead to measurable physiological responses (cf. Russo, Vempala, & Sandstrom, 2013), which may be mediated by preference and pleasure. Although recognized to be of substantial importance in music listening, there is only a limited amount of research into the effects of music preference. In the current study, to prevent effects of agency and being in control, stimuli were preselected rather than self-selected by participants, and liking, familiarity, induced and perceived emotion ratings were collected for stimulus pairs that differed highly in genre, arousal and valence. The resulting difference scores are used to assess whether changes in the different measures would reveal specific effects of differences in liking or arousal on cognitive performance.

II. METHODS

Although the project is currently still ongoing, the reported results are based on the first 39 full data sets (of 64 needed for full counterbalancing of stimuli). Nonmusician participants were recruited from the Leiden student body, with 29 females, 10 males and a mean age of 21 years old (SD=0.8 years). In separate sessions with continuous heart rate measurement, participants each heard two out of eight musical stimuli, counterbalanced for order and arousal/valence content, lasting around 5 minutes. After rating their liking of the music as well as induced and perceived arousal and valence using the Self-Assessment Manikin (SAM, Bradley & Lang, 1994), they completed a mental rotation task (MRT, Vandenberg & Kuse, 1978). Heart rate was measured using a Biopac electrocardiogram (ECG) setup, and analysed in Matlab 2012b. The experiment was run in E-Prime 2.0, and the data analysed in SPSS 23.0 (IBM).

III. RESULTS

There were no significant correlations between the difference scores of the MRT and those of any of the other measures (familiarity, liking, induced arousal, induced valence, intended arousal, intended valence, heart rate or heart rate variability during music, and heart rate or heart rate variability during MRT). Thus it appears that MRT performance is not related to liking ($r=-0.057$, $p=0.728$), or to induced, perceived or physiological arousal ($r=-0.247$, $p=0.13$; $r=0.038$, $p=0.817$; and $r=-0.1$, $p=0.566$, respectively). Differences in liking are highly related to differences in familiarity ($r=0.70$, $p<0.001$) and induced emotion ($r=0.47$, $p=0.002$), and differences in self-reported induced arousal are related to differences in heart rate during the cognitive task ($r=0.38$, $p=0.025$), supporting the current design and stimuli that were used.

IV. CONCLUSION

Although the results are still preliminary, our data suggest that even if cognitive performance increases after music listening, which is not tested in the current experiment, the effects do not appear to be mediated by either music preference or arousal. If this pattern persists as data collection continues, there is ground to search for other explanations of why cognitive performance should increase after music listening. Given the implications of effects of music listening on cognition, especially for clinical groups such as dementia patients, it is important to unravel the mechanism behind responses to music when applied to cognitive functioning.

REFERENCES

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25*, 49–59.

Rauscher, F. H., Shaw, G. L., & Ky, K. N. (1993). Music and spatial task performance. *Nature, 365*, 611.

Russo, F,A., Vempala, N, N., Sandstrom, G, M., (2013). Predicting musically induced emotions from physiological inputs: linear and neural network models *Front. Psychol*, *4,* 468.

Schellenberg G, E. (2014). Music in the social and behavioral sciences: an encyclopedia. *Sage editors, 4*, 717-718.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 47*, 599–604.

Music training, empathy, and theory of mind

Cara Z. Fesjian, Andrew Tung, Hwanjun Chung, Ryan Veiga, Assal Habibi

***Abstract*—** Music training has been associated with cognitive benefits such as language development and academic success, but the degree to which music training can influence development of social and emotional skills such as empathy remains unclear. Given that music is often considered to be intimately connected with emotions and feelings, it is reasonable to expect that music training can influence development of socio-emotional skills as well as cognitive abilities. This study aimed to investigate whether music training during childhood is associated with development of cognitive and affective empathy in adulthood and whether the age at which music training begins along with duration of practice would mediate the association between music training and empathy skills. Three groups of participants (early onset musicians, late onset musicians, and nonmusicians) completed an online survey that incorporated three assessments of empathy: Reading the Mind in the Eyes Test, Reading the Mind in the Voice Test, and the Interpersonal Reactivity Index. The results show that musicians with an early onset age of training (age five or below) performed significantly better on the Reading the Mind in the Eyes Test compared to late onset musicians and nonmusicians. There were no other significant differences between the three groups. The findings suggest that music training, in particular when it begins in early childhood, can influence development of cognitive empathy skills. Other factors such as cognitive capacities, personality traits, and types of music training (group versus individual) should be considered in interpreting the findings.

***Keywords*—**Empathy, music training, onset age, theory of mind.

I. Introduction

PLAYING a musical instrument is a complex task that of necessity engages many different brain regions, as it requires the concurrent recruitment of sensory, motor, executive, and affective systems. The mastering of this rich and demanding process requires regular and intense practice, and it involves motor coordination as well as emotional communication with other players and listeners. The combination of such demands is likely to influence development of cognitive and socio-emotional skills.

Accordingly, several investigators have reported differences in cognitive abilities, such as language development [1]-[3], visuospatial skills [4], [5], reading skills [6]-[10], general intelligence [11]-[14], and academic achievement [15], [16], as well as structural brain differences [17]-[19], between musicians and nonmusicians. Although the evidence for cognitive benefits of music training is strong, the extent to which benefits for non-musical abilities extend to emotional and social skills remains unclear. Music is intimately connected with emotion and feelings [20]-[22] and can evoke a broad range of emotions, from joy and peacefulness to sadness and fear [23]. Music also has an enormous ability to connect people with each other, though the neural correlates of this connection are not completely understood [24], [25]. Therefore, it is reasonable to expect that involvement in music can improve emotional and social abilities, although to date the available evidence is inconclusive.

Schellenberg showed that adults with at least eight years of private music lessons, despite having a higher IQ compared to their untrained counterparts, did not show an advantage in overall emotional intelligence [13]. Similarly, in a study with a large sample of elementary school children, music training was shown to be positively associated with IQ and academic achievement but not with social skills [16]. Finally, in a group of elementary school students who were assigned to one year of drama, keyboard, or voice lessons, only the drama group showed significant improvements in social skills [26]. These findings seem to suggest that non-musical associations with music training may be limited to measures of cognitive abilities. However, others have shown that when empathy was measured in children after one year of musical group interaction, the music group showed higher empathy scores than a control group [27]. Similarly, Kirschner and Tomasello showed that cooperative and helpful behavior was enhanced in four-year-olds following joint music-making [28], and school children who participated in a musical program with a specific focus on empathy — through singing or composing songs — showed higher empathy levels when compared to same-age children who did not participate in the program [29]. Empathy has also been associated with both the potential to feel sadness from sad music [30] and the potential to like listening to sad music [31]. Furthermore, higher levels of empathy yield better identification of musicians' intentions during performance [32], and empathy has been shown to mediate the degree to which people can perceive emotion from music [33].

One important factor in assessing the relationship between music training and development of socio-emotional skills is one's onset age of music training, given the important association between onset age of music training and changes of brain and behavior during development. Improvements in auditory processing, such as pitch and timbre processing, have been shown to relate to sound input experienced at a young age [34]. Onset age of music training has also been shown to correlate with anatomical differences specifically in the motor

Assal Habibi is with the Brain & Creativity Institute, University of Southern California, Los Angeles, CA 90089 (phone: 213-740-3588; e-mail: assal.habibi@usc.edu).

Cara Z. Fesjian was with the Thornton School of Music and Dornsife College of Letters, Arts, & Sciences, University of Southern California, Los Angeles, CA 90089. She is now with the Brain & Creativity Institute, University of Southern California, Los Angeles, CA 90089 (e-mail: fesjian@usc.edu).

cortex [35] and corpus callosum [36], [37]. There is also strong evidence suggesting that the acquisition of absolute pitch and its persistence into adulthood depends on a critical period that ends around age six [38] and that the probability of acquiring absolute pitch is related to onset age [39], [40]. Further, onset age of music training positively influences performance on visual- and auditory-motor synchronization tasks [41]-[43]. Given these relationships, it is plausible that musicians who begin studying music during a sensitive period show increased levels of empathy in adulthood.

The aim for this study was to investigate whether onset age of music training during childhood can impact development of socio-emotional skills, including theory of mind and empathy. We hypothesized that musicians with an early onset age of music training would outperform musicians with late onset age of training as well as nonmusicians in both measures of empathy and theory of mind.

II. Methods

A. Participants

One-hundred-thirty-six adults took part in our study via an online survey distributed through the University of Southern California and social media platforms. Study protocols were approved by the University of Southern California Institutional Review Board. Twelve participants did not complete the measures in full and were removed from the analysis. Following data collection, a skewed sample of participants below the age of 18 and above the age of 35 was observed, and therefore the sample was adjusted to remove all subjects not within the 18 to 35 age bounds. Given the nature of the online survey and potential for invalid results, all data from subjects who had scores two standard deviations or more away from the group mean on any test were also eliminated. In all, 91 participants total (57 female, Mean age = 22.99, SD = 4.13) were included in the analysis and divided into three groups a priori: early onset musicians (n = 18, 14 female, Mean age = 24.39, SD = 5.55), late onset musicians (n = 34, 20 female, Mean age = 23.06, SD = 4.03), and nonmusicians (n = 39, 23 female, Mean age = 22.28, SD = 3.32). We specifically defined early onset musicians as individuals who began training at or below the age of five, and late onset musicians as individuals who began training after the age of five. Both early and late onset musicians had at least seven years of music training, and private lessons or ensemble training qualified as music training. Nonmusicians had no music training. Each subject received a $5 Amazon online gift card upon completion of the study as compensation for participation.

B. Materials

Participants were assessed using a survey (Qualtrics-based) that consisted of four sections: background questionnaire, Reading the Mind in the Eyes Test (RMET) [44], Reading the Mind in the Voice Test (Voice Test) [45], and the Interpersonal Reactivity Index (IRI) [46]. Background

TABLE I
PARTICIPANT DEMOGRAPHICS

	Musicians (n = 52) Early onset (n = 18)	Musicians (n = 52) Late onset (n = 34)	Nonmusicians (n = 39)
Mean age (SD)	24.39 (5.55)	23.06 (4.03)	22.28 (3.32)
Females	14	20	23
Males	4	14	16
Mean onset age of training (SD)	4.67 (.49)	9.27 (1.97)	
Mean years of training (SD)	19.72 (5.42)	13.79 (4.63)	
Instrument			
Piano	12	9	
Strings	2	7	
Winds	0	7	
Brass	1	4	
Voice	3	5	
Percussion	0	2	
Ensemble?			
Yes	8	25	
No	10	9	

Descriptive statistics for age and sex of all participants; descriptive statistics for onset age, years of training, primary instrument, and current ensemble involvement of musicians.

information included age, sex, and questions regarding the duration and details of one's music training. Music interest, genre preference, and total music listening in hours per week were also assessed.

The RMET is compiled of 36 images of people's eyes that participants are to assign to the correct emotion. For each image, participants are given four emotions and are asked to choose the emotion that best fits the eyes. It has been validated as a measure to detect individual differences in social sensitivity and adult "mentalizing" [44]. Similarly, the Voice Test is compiled of 25 voice recordings that participants are supposed to assign to the correct emotion, given four options. It detects individual differences in empathy in regards to auditory stimuli [45]. Both are considered tests of cognitive empathy [44], [45].

The IRI is an assessment of empathy in adults involving four subscales and their relationships with measures of social functioning, self-esteem, emotionality, and sensitivity [46]. Each scale measures a distinct component of empathy. Perspective Taking measures the degree to which one is able to take the perspective of another; Fantasy measures one's potential for emotional identification with characters in books, films, and other media; Empathic Concern measures the degree of feeling emotional concern for others; and Personal Distress measures one's negative feelings in response to the distress of others. Each given statement in the IRI pertains to one of the four scales, and participants must rank each statement depending on how well it describes them.

C. Procedure

Participants consented to participate in the study at the beginning of the survey. They were asked to take the survey in a quiet environment, with headphones, to eliminate interference during the Voice Test. Given that participants were recruited online, a uniform testing environment was not conceivable. Participants first completed the background questionnaire, followed by the IRI, RMET, and the Voice Test. The order of the tests was not counterbalanced across participants.

D. Analysis

Analysis of variance revealed no differences in age, F (2, 88) = 1.63, $p = .20$, or sex, χ^2 (2) = 2.20, $p = .33$, between the three groups, and therefore these factors were not included in any subsequent analysis. We then analyzed the age-matched sample data with a series of univariate ANOVAs, with group (early onset, late onset, and nonmusicians) as the independent factor. Pearson correlations were used to examine relationships between onset age and years of training and all tests in the battery. For all analyses, an alpha level of .05 was maintained. All statistical analyses were performed using SPSSStatistics (v. 21).

III. Results

A summary of participant demographics, instruments, and current ensemble involvement is provided in Table 1. There was a significant difference between the three groups in scores on the RMET, F (2, 88) = 4.70, $p = .01$, $\eta_p^2 = .096$. Post-hoc contrasts revealed a significant difference between early onset musicians and late onset musicians, $p = .004$, as well as a significant difference between early onset musicians and nonmusicians, $p = .01$, in scores on the RMET.

There was no significant difference between the three groups in scores on the Voice Test, F (2, 88) = 1.90, $p = .16$, although the early onset musicians performed better compared to the other two groups. There were also no significant differences between the three groups in scores of Perspective Taking, F (2, 88) = 2.00, $p = .14$, Fantasy, F (2, 88) = .24, $p = .79$, Empathic Concern, F (2, 88) = 1.58, $p = .21$, or Personal Distress, F (2, 88) = 1.51, $p = .23$, on the IRI. A summary of performance for the three groups on all tests is provided in Table 2.

There was also a significant negative correlation between onset age of music training and scores on the RMET, $r = -.27$, $n = 52$, $p = .05$. There were no further relationships between onset age or years of training and performance on the measures.

IV. Discussion

This study aimed to explore the relationship between music training, empathy, and theory of mind. Few studies have explored the extent to which empathy is improved by music practice in the long term. Given influences of music on emotion and the various reported cognitive benefits of music training especially from a young age, we hypothesized that early onset musicians may have enhanced socio-emotional skills as a result of their training.

When comparing the three groups (early onset musicians, late onset musicians, and nonmusicians), we found that early onset musicians performed significantly better than late onset musicians and nonmusicians on the RMET. Scores on the RMET were further correlated with the age at which musicians began training, but not with their number of years of training.

Both the group difference and the correlation provide support for a potential lasting enhancement of cognitive empathy with early onset music training. Given that musicians are taught to convey emotions through performance, and early onset musicians have had longer exposure to this training, they may have enhanced abilities in recognizing emotional states of others. Further, typically developing children's abilities to conceptualize theory of mind develops by age five [47], [48]. The convergence of these musical and empathic periods in development, coupled with the plasticity of a child's brain, may promote a mechanism for the enhancement of empathy in the long term. The nature of early music training methods should also be considered in interpreting these results. Early childhood music training often focuses on group musicianship lessons and exercises that inherently involve social interaction thereby providing opportunities for socio-emotional benefits. Enhanced cognitive abilities and their association with early onset music training are another possible contributing factor. However, information regarding cognitive capacities, such as IQ, was not collected in our study and thus was not controlled for. We also did not have any information regarding other extracurricular activities during childhood from this population.

In relation to the performance on the Voice Test, given that adult musicians as well as six-year-olds with one year of music training are better at identifying emotions in speech based on prosody [49], it is surprising that musicians did not perform significantly better on the Voice Test compared to nonmusicians. It is possible that emotional intelligence is the main contributor for better emotion recognition in prosody, not music training per se [50]. Finally, given the nature of the online survey and the lack of a uniform testing environment, there may have been sound interference that influenced scores on the Voice Test.

Our musician sample was compiled of a wide range of instrumentalists: some played in ensembles, and others took private lessons and played solo. On the one hand, restricting our musician population to a certain group of players (e.g., only pianists, who are used to practicing and playing alone)

TABLE II
PERFORMANCE ON EMPATHY MEASURES

Test	Mean (SD) Early onset	Late onset	Nonmusicians
RMET	30.61 (2.45)	28.06 (3.60)	28.44 (2.52)
Voice Test	19.06 (2.90)	17.88 (2.66)	18.72 (1.57)
Perspective-Taking	17.61 (4.33)	18.00 (3.11)	19.64 (4.99)
Fantasy	19.17 (5.87)	19.32 (5.12)	20.03 (5.05)
Empathic Concern	21.94 (3.80)	21.06 (3.43)	19.85 (5.26)
Personal Distress	12.39 (4.10)	10.00 (5.26)	10.87 (4.46)

Participant scores on all empathy measures. Early onset musicians, $n = 18$; Late onset musicians, $n = 34$; Nonmusicians, $n = 39$. RMET = Reading the Mind in the Eyes Test; Voice Test = Reading the Mind in the Voice Test; Perspective-Taking, Fantasy, Empathic Concern, & Personal Distress subscales of the Interpersonal Reactivity Index (IRI).

would have given us more specific data. On the other hand, gathering data from a group of musicians with different specifications provides us with a broader look at the differences between musicians and nonmusicians. Nevertheless, future research should isolate various music training environments in order to investigate the potentially different influences ensemble training has on empathy compared to solo training.

The present study provides evidence for the possibility of a sensitive period for development of associations of music training and empathy skills. These findings support previous work that shows associations between music training and emotion recognition skills [27], [51]. While the results from this study deepen our understanding of the relationship between music training and empathic abilities, future research should look towards a predictive model that would take a variety of training methods, ensemble involvement, environmental and innate factors into account.

References

[1] J. Chobert, C. Francois, J. Velay, and M. Besson, "Twelve months of active musical training in 8- to 10-year-old children enhances the preattentive processing of syllabic duration and voice onset time," *Cerebral Cortex*, vol. 24, no. 4, pp. 956–967, Apr. 2014.

[2] C. Francois, J. Chobert, M. Besson, and D. Schon, "Music training for the development of speech segmentation," *Cerebral Cortex*, vol. 23, no. 9, pp. 2038–2043, Sep. 2013.

[3] D. L. Strait and N. Kraus, "Biological impact of auditory expertise across the life span: Musicians as a model of auditory learning," *Hearing Research*, vol. 308, pp. 109–121, Feb. 2014.

[4] R. Brochard, A. Dufour, and O. Després, "Effect of musical expertise on visuospatial abilities: Evidence from reaction times and mental imagery," *Brain and Cognition*, vol. 54, no. 2, pp. 103–109, Mar. 2004.

[5] V. Sluming, J. Brooks, M. Howard, J. J. Downes, and N. Roberts, "Broca's area supports enhanced Visuospatial Cognition in orchestral musicians," *Journal of Neuroscience*, vol. 27, no. 14, pp. 3799–3806, Apr. 2007.

[6] S. Moreno, C. Marques, A. Santos, M. Santos, S. L. Castro, and M. Besson, "Musical training influences linguistic abilities in 8-year-old children: More evidence for brain plasticity," *Cerebral Cortex*, vol. 19, no. 3, pp. 712–723, Mar. 2009.

[7] S. Moreno, D. Friesen, and E. Bialystok, "Effect of music training on promoting preliteracy skills: Preliminary causal evidence," *Music Perception: An Interdisciplinary Journal*, vol. 29, no. 2, pp. 165–172, Dec. 2011.

[8] F. Degé and G. Schwarzer, "The effect of a music program on phonological awareness in preschoolers," *Frontiers in Psychology*, vol. 2, no. 24, pp. 7-13, Jun. 2011.

[9] J. M. Piro and C. Ortiz, "The effect of piano lessons on the vocabulary and verbal sequencing skills of primary grade students," *Psychology of Music*, vol. 37, no. 3, pp. 325–347, Jul. 2009.

[10] A. Tierney and N. Kraus, "Music training for the development of reading skills," in *Progress in Brain Research*, vol. 207, M. M. Merzenich, M. Nahum, T. M. Van Vleet, Eds. Burlington: Academic Press, 2013, pp. 209-241.

[11] E. Glenn Schellenberg, "Music lessons enhance IQ," *Psychological Science*, vol. 15, no. 8, pp. 511–514, Aug. 2004.

[12] E. G. Schellenberg, "Music and cognitive abilities," *Current Directions in Psychological Science*, vol. 14, no. 6, pp. 317–320, Dec. 2005.

[13] E. Glenn Schellenberg, "Music lessons, emotional intelligence, and IQ," *Music Perception: An Interdisciplinary Journal*, vol. 29, no. 2, pp. 185–194, Dec. 2011.

[14] E. G. Schellenberg, "Examining the association between music lessons and intelligence," *British Journal of Psychology*, vol. 102, no. 3, pp. 283–302, Aug. 2011.

[15] K. R. Fitzpatrick, "The effect of instrumental music participation and socioeconomic status on Ohio fourth-, sixth-, and ninth-grade proficiency test performance," *Journal of Research in Music Education*, vol. 54, no. 1, pp. 73–84, Apr. 2006.

[16] E. G. Schellenberg, "Long-term positive associations between music lessons and IQ," *Journal of Educational Psychology*, vol. 98, no. 2, pp. 457–468, May 2006.

[17] V. Sluming, T. Barrick, M. Howard, E. Cezayirli, A. Mayes, and N. Roberts, "Voxel-Based Morphometry reveals increased gray matter density in Broca's area in male symphony orchestra musicians," *NeuroImage*, vol. 17, no. 3, pp. 1613–1622, Nov. 2002.

[18] C. Gaser and G. Schlaug, "Brain structures differ between musicians and non-musicians," *Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, vol. 23, no. 27, pp. 9240–9245, Oct. 2003.

[19] G. Schlaug, "Effects of music training on the child's brain and cognitive development," *Annals of the New York Academy of Sciences*, vol. 1060, no. 1, pp. 219–230, Dec. 2005.

[20] A. Habibi and A. Damasio, "Music, feelings, and the human brain," *Psychomusicology: Music, Mind, and Brain*, vol. 24, no. 1, pp. 92–102, 2014.

[21] P. G. Hunter and E. G. Schellenberg, "Music and emotion" in *Music perception.* New York: Springer, 2010, pp. 129-164.

[22] P. N. Juslin, S. Liljeström, D. Västfjäll, G. Barradas, and A. Silva, "An experience sampling study of emotional reactions to music: Listener, music, and situation," *Emotion*, vol. 8, no. 5, pp. 668–683, Oct. 2008.

[23] M. Zentner, D. Grandjean, and K. R. Scherer, "Emotions evoked by the sound of music: Characterization, classification, and measurement," *Emotion*, vol. 8, no. 4, pp. 494–521, Aug. 2008.

[24] T. DeNora, "Emotion as social emergence: Perspectives from music sociology," in *Handbook of music and emotion: Theory, research, applications,* P. N. Juslin, and J. A. Sloboda, Eds. Oxford: Oxford University Press, 2010, pp. 159–183.

[25] P. N. Juslin and R. Timmers, "Expression and communication of emotion in music performance," in *Handbook of music and emotion: Theory, research, applications*, P. N. Juslin, and J. A. Sloboda, Eds. Oxford: Oxford University Press, 2010, pp. 453-489.

[26] T. R. Goldstein and E. Winner, "Enhancing empathy and theory of mind," *Journal of Cognition and Development*, vol. 13, no. 1, pp. 19–37, Jan. 2012.

[27] T. Rabinowitch, I. Cross, and P. Burnard, "Long-term musical group interaction has a positive influence on empathy in children," *Psychology of Music*, vol. 41, no. 4, pp. 484–498, Jul. 2013.

[28] S. Kirschner and M. Tomasello, "Joint music making promotes prosocial behavior in 4-year-old children," *Evolution and Human Behavior*, vol. 31, no. 5, pp. 354–364, Sep. 2010.

[29] F. Laurence, "Music and empathy," in *Music and conflict transformation: Harmonies and dissonances in geopolitics*, O. Urbain, Ed. London: I.B Tauris, 2008, pp. 13-25.

[30] J. K. Vuoskoski and T. Eerola, "Can sad music really make you sad? Indirect measures of affective states induced by music and autobiographical memories," *Psychology of Aesthetics, Creativity, and the Arts*, vol. 6, no. 3, pp. 204–213, Aug. 2012.

[31] J. K. Vuoskoski and W. F. Thompson, "Who enjoys listening to sad music and why?," *Music Perception: An Interdisciplinary Journal*, vol. 29, no. 3, pp. 311–317, Feb. 2012.

[32] C. Wöllner, "Is empathy related to the perception of emotional expression in music? A multimodal time-series analysis," *Psychology of Aesthetics, Creativity, and the Arts*, vol. 6, no. 3, pp. 214–223, Aug. 2012.

[33] H. Egermann and S. McAdams, "Empathy and emotional contagion as a link between recognized and felt emotions in music listening," *Music Perception: An Interdisciplinary Journal*, vol. 31, no. 2, pp. 139–156, Dec. 2013.

[34] L. J. Trainor, "Are there critical periods for musical development?," *Developmental Psychobiology*, vol. 46, no. 3, pp. 262–278, Apr. 2005.

[35] K. Amunts *et al.*, "Motor cortex and hand motor skills: Structural compliance in the human brain," *Human Brain Mapping*, vol. 5, no. 3, pp. 206–215, Jan. 1997.

[36] G. Schlaug, "Increased corpus callosum size in musicians," *Neuropsychologia*, vol. 33, no. 8, pp. 1047–1055, Aug. 1995.

[37] C. J. Steele, J. A. Bailey, R. J. Zatorre, and V. B. Penhune, "Early musical training and white-matter plasticity in the corpus callosum: Evidence for a sensitive period," *Journal of Neuroscience*, vol. 33, no. 3, pp. 1282–1290, Jan. 2013.

[38] H. Takeuchi and S. H. Hulse, "Absolute pitch," *Psychological Bulletin*, vol. 113, no. 2, pp. 345–361, Mar. 1993.

[39] S. Baharloo, P. A. Johnston, S. K. Service, J. Gitschier, and N. B. Freimer, "Absolute pitch: An approach for identification of genetic and nongenetic components," *The American Journal of Human Genetics*, vol. 62, no. 2, pp. 224–231, Feb. 1998.

[40] D. Deutsch, T. Henthorn, E. Marvin, and H. Xu, "Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a speech-related critical period," *The Journal of the Acoustical Society of America*, vol. 119, no. 2, pp. 719–722, Feb. 2006.

[41] J. A. Bailey and V. B. Penhune, "Rhythm synchronization performance and auditory working memory in early- and late-trained musicians," *Experimental Brain Research*, vol. 204, no. 1, pp. 91–101, Jul. 2010.

[42] J. A. Bailey and V. B. Penhune, "A sensitive period for musical training: Contributions of age of onset and cognitive abilities," *Annals of the New York Academy of Sciences*, vol. 1252, no. 1, pp. 163–170, Apr. 2012.

[43] D. Watanabe, T. Savion-Lemieux, and V. B. Penhune, "The effect of early musical training on adult motor performance: Evidence for a sensitive period in motor learning," *Experimental Brain Research*, vol. 176, no. 2, pp. 332–340, Jan. 2007.

[44] S. Baron-Cohen, S. Wheelwright, J. Hill, Y. Raste, and I. Plumb, "The 'Reading the mind in the Eyes' test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism," *Journal of Child Psychology and Psychiatry*, vol. 42, no. 2, pp. 241–251, Feb. 2001.

[45] O. Golan, S. Baron-Cohen, J. J. Hill, and M. D. Rutherford, "The 'Reading the mind in the Voice' test-revised: A study of complex emotion recognition in adults with and without autism spectrum conditions," *Journal of Autism and Developmental Disorders*, vol. 37, no. 6, pp. 1096–1106, Jul. 2007.

[46] M. H. Davis, "Measuring individual differences in empathy: Evidence for a multidimensional approach," *Journal of Personality and Social Psychology*, vol. 44, no. 1, pp. 113–126, Jan. 1983.

[47] H. M. Wellman, *The child's theory of mind*. Cambridge, MA: MIT Press, 1990.

[48] H. M. Wellman and K. Bartsch, "Before belief: Children's early psychological theory," in *Children's Early Understanding of Mind: Origins and Development*, C. Lewis, P. Mitchell, Eds. London: Psychology Press, 1994, pp. 331-354.

[49] W. F. Thompson, E. G. Schellenberg, and G. Husain, "Decoding speech prosody: Do music lessons help?" *Emotion*, vol. 4, no. 1, pp. 46–64, Mar. 2004.

[50] C. G. Trimmer and L. L. Cuddy, "Emotional intelligence, not music training, predicts recognition of emotional speech prosody," *Emotion*, vol. 8, no. 6, pp. 838–849, Dec. 2008.

[51] C. E. Parsons, K. S. Young, E. M. Jegindø,, P. Vuust, A. Stein, and M. L. Kringelbach, "Music training and empathy positively impact adults' sensitivity to infant distress," *Frontiers in Psychology*, vol. 5, Dec. 2014.

Autism symptomatology in the general population and processing of musical emotion and structure

Eve-Marie Quintin[1], Julia Donahue[1], and Sophia Metcalf[1]
[1] Behavior, Autism, and NeuroDevelopment (BAND) research group, Educational and Counseling Psychology Department, McGill University, Montreal, Quebec, Canada

There is growing evidence that musical perception and ability are strengths of individuals with autism spectrum disorder (ASD). To date, most studies have examined how individuals with ASD process either the emotional or cognitive aspects of music. Thus, the question of what draws individuals with ASD to music remains unresolved. The current study aims to assess whether (sub-clinical) autism symptomatology in the general population impacts the relationship between processing of musical emotions and structure. We propose a dual experimental task using a common set of musical stimuli where participants must 1. select which emotion (happy, sad, or scary; forced choice paradigm) best describes the stimuli and 2. detect chromatic or diatonic variations in the structure of the stimuli (same-different paradigm). Higher autism symptomatology, as measured with the general population Adult Autism Spectrum Quotient and Social Responsiveness Scale, should be associated with greater discrepancy in reaction time and total score between musical task 1 (emotion selection) and 2 (structure detection)*. NEO personality scores will also be correlated with the discrepancy between task 1 and 2. Musical interest and experience will be controlled for using the Goldsmiths Musical Sophistication Index. Findings will be discussed within the context of Baron-Cohen's empathizing-systemizing theory, which suggests that discrepancies between socio-emotional and cognitive profiles are distributed along a continuum in the general population, with ASD representing one extreme of this continuum.
*Data collection is ongoing and should be completed before the conference

How do non-musicians communicate musical emotions? Expressive timing in a self-pacing paradigm

Haley E. Kragness[1] and Laurel J. Trainor[1,2,3]
[1]Psychology, Neuroscience and Behaviour, McMaster University, Hamilton, ON, Canada
[2]McMaster Institute for Music and the Mind, McMaster University, Hamilton, ON, Canada
[3]Rotman Research Institute, Baycrest Hospital, Toronto, Canada

In the research reported here we investigate expressive timing in the performances of non-musicians using a new methodology. A number of studies have demonstrated that listeners can reliably detect simple emotions in performances by trained musicians. This body of work has led to the hypothesis that there may be a "shared code" for emotional prosody in music and other domains, such as language. For obvious reasons, previous research has been limited to examining performances by trained musicians, so the contribution of formal music training to expressive ability is unclear. However, the popularity of music-making games and viral smartphone applications (such as Guitar Hero and Virtual Piano) suggests that even non-musicians will readily engage in producing music when it is accessible. In the present experiment, we invite musically trained and untrained participants to perform musical passages with different emotions using a self-pacing paradigm. First, participants passively listen to a short chord sequence. Next, they are asked to self-pace through the sequence by pressing a computer key to control the onset and offset of each chord in succession, with the goal of conveying a particular emotion (happy, sad, angry, or peaceful) to a second participant. We measure participants' chosen *tempo* (speed), *rubato* (inter-onset-interval variability), and *articulation* (ratio of filled to empty space). Data collection is ongoing. Results of this experiment will examine how musically untrained participants use temporal cues to communicate emotions in music. We hypothesize non-musicians will use similar timing patterns to convey emotions as musicians. Future work will use "performances" collected in the current experiment to investigate whether people can correctly identify the intended emotions. This simple paradigm could also be used to examine the expressive musical production of young children.

Listener-normalized musical affect: conceptualizing "felt" musical emotions

Joseph Plazak[*1]

[*]School of Music, Illinois Wesleyan University, U.S.A
[1]jplazak@iwu.edu

ABSTRACT

One commonly addressed distinction within the literature on musical affect is the difference between perceived and felt emotional experiences. This paper presents a conceptual framework for understanding how felt musical emotions differ from perceived musical emotions. Theories of perceived musical emotion often assume that musical sounds are normalized relative to their sound source. In this way, perceived emotions are often "externally" normalized relative to the capacities of a given instrument or ensemble. Felt musical emotions are, by definition, "internal" to the listener, and therefore, our understanding of these emotions might benefit from a process of "listener-normalization." By employing "psychomechanical" attributes of sound, such as perceived sound source size and/or perceived sound source energy, it is possible to normalize sounds relative to the receiver/listener. When listeners normalize sounds relative to themselves, they may be able to determine their ability to "cope" with the size and energy of the sound source, and thereafter, activate proper "felt" emotional responses. This paper addresses current issues relating to our understanding of induced musical emotion, and in particular, details how the process of listener-normalization might explain some variability within experimental research on "felt" musical emotions.

I. INTRODUCTION

There are a number of complexities concerning the ability of music to communicate emotional information, but one commonly made distinction pertains to the difference between "perceived" and "felt" affective communication. Since Gabrielsson (2002) reinvigorated the question of "internal locus of emotion" vs. "external locus of emotion" within musical communication, many studies have investigated this distinction (Schubert, 2013). The purpose of the present paper is to discuss how the process of listener-normalization might explain some variability within experimental research on "felt" musical emotion, and to encourage further research via this particular conceptualization.

II. SOURCE NORMALIZATION: PERCEIVED EMOTIONS

Perceived/expressed musical emotions are those that can successfully communicate emotional information to a receiver/listener, without necessarily inducing these emotions. Normalization plays an important role in the successful communication of such perceived emotions. For example, when perceiving a sad musical passage performed on a given instrument, such as the violin, one might claim that the music sounds sad because it is low in pitch, utilizes a dark timbre, and/or is generally quiet throughout (see Juslin & Timmers, 2010 for a summary of communicative codes in music). It is important to note that such claims typically assume the normalization and relativity of musical cues. In the case of the sad violin example, the characteristics of sad music are normalized relative to the capacities of the violin, and thus we might more accurately claim that a given passage is relatively low in pitch "for a violin," relatively dark in timbre "for a violin," and relatively quiet "for a violin." By definition, perceived emotions (with the possible exception of self-perception) are heard relative to an "external" sound source, and therefore, sound-source normalization is useful for perceiving variations of expressive intent.

Within the literature on affective speech, it is common to explicitly normalize cues relative to the sound source. For example, in a study that analyzed the acoustic features of confidence in speech, Jiang & Pell (2015) normalized the f_0 contours of each stimulus relative to the mean of each speaker's minimum frequency that was spoken within a neutral statement. In studies on pitch perception in music, sound-source normalization is sometimes ignored, thus leading to puzzling associations across the literature. For example, Juslin & Timmer's (2010) review of associations between affect and musical structures revealed that "High Pitch Level" has been previously associated with a diverse range of perceived emotional information, including: happiness, pleading, anger, and fear. With such a wide range of pitch associations, a standardized source-normalization procedure, much like those found in the affective speech literature, might be beneficial for understanding how the abovementioned associations differ. By normalizing musical pitch relative to a sound source, we might more accurately capture interactions between pitch and timbre (i.e. register); an interaction known to result in different perceived emotional information (Huron, et al., 2006).

III. LISTENER NORMALIZATION: FELT EMOTIONS

When discussing felt/induced musical emotion, the role of normalization might be considered more abstract. Because felt emotional responses are "internal" by definition, it might be useful to consider normalizing felt emotional responses relative to the receiver (i.e. listener) rather than the sound source.

Normalizing a set of musical cues against a listener might initially seem problematic. While we can think of an instrument as being relatively low in pitch or dark in timbre, we typically do not describe ourselves in this same way. Most musical descriptors are not particularly apt for also describing listeners. Therefore, in order to normalize musical sounds relative to the listener, we will need to examine a unique set of sound descriptors.

Beyond acoustic and musical descriptors of sound, psychophysical and psychomechanical descriptors of sound (McAdams et. al, 1994; Stoelinga, 2009) may prove to be useful for normalizing sound stimuli relative to a listener. Psychomechanical attributes of sound include characteristics that can be inferred about a source from sound cues alone, such as perceived sound source size, perceived sound source energy, sound source material, sound source proximity, etc. One unique feature of these cues is that they are common to both the sound source and the listener; both the source and the receiver will have a physical size, amount of energy, physical location, etc. For example, via psychomechanical sound descriptors, it is possible to assess if a given sound source is much smaller or much larger than a particular listener, as might be the case when comparing performances of ukulele music with pipe organ music. The former sound source is likely to be relatively smaller than the listener, whereas the latter sound source is likely to be much bigger. A similar comparison could be made between an energetic sound source and a lethargic sound receiver.

The conceptual framework brought forward here is that induced musical emotions stem from cues that are normalized relative to a given musical receiver (i.e. listener), and perceived musical emotions stem from cues that are normalized to a given musical sender (i.e. instrument or ensemble). In many ways, this idea represents what is already inherent within the connotative meaning of the words "perceived" and "felt." Perceived emotion, by definition, implies a sense of outwardness, especially with regards to perceiving emotion within music, whereas felt emotion, on the other hand, implies a sense of inwardness. Therefore, the process of normalization may play an important role towards understanding the difference between these two emotional loci.

IV. DOUBLE-NORMALIZATION

In order to illustrate the utility of "double-normalizing" affective musical communication, that is, normalizing cues relative to both the sound source (perceived emotion) and the sound receiver (felt emotion), consider the various ways in which one might respond to the sound of a human vocalizing an angry growl.

Being able to perceive that a sound source is expressing anger involves multiple cues, such as sharp amplitude envelopes (Juslin, 1997), lower f_0 (Morton, 1977), large f_0 range (Williams & Stevens, 1972), etc. Recall that these cues are not absolute, but instead perceived relative to a given sound source. Without the ability to normalize the pitch and intensity ranges of the sound source, it might not be possible to understand how we perceive expressive intent. Normalization facilitates the ability to hear sounds as relatively low and/or relatively sharp, and thereafter "perceive" that a given sound source may be angry.

Beyond sound source-normalization, listener-normalization may provide insight into plausible induced emotional responses to this angry vocalization. For simplicity, a singular psychomechanical property will be used in this example: sound source size. Assume that the listener not only perceives that the sound source is angry (via the cues mentioned above), but also simultaneously perceives the physical size of the sound source (see Patterson, et al. 2008 for an explanation of auditory size perception). Here, we will examine three source size exemplars: same source size (an angry adult), much smaller source size (an angry dwarf), and much larger source size (an angry 10-foot tall giant). In all three cases, the "perceived" emotional message is the same (i.e. anger), yet through the process of listener-normalization, the "felt" emotional response to the angry vocalization could be manifested in many different ways.

Hypothetically, if an angry vocalization was produced by a sound source of approximately the same size as the listener, the felt affective responses might be expected to vary widely, ranging from indifference to fear. In the second example, in which the angry vocalization was produced from a much smaller dwarf-like sound source, different induced responses might be expected, such as laughter, ridicule, or compassion. In the final example, in which the angry vocalization was produced by a much larger giant-like sound source, one might expect induced responses such as terror or panic. These hypothetical responses are closely related to the sound receiver's ability to cope with the angry sound source.

All of the potential induced responses in the example above are theoretically derived from normalizing a single psychomechanical source cue relative to the listener. However, many concurrent psychomechanical cues should be expected to play a role in induced emotional responses (e.g. source energy, source proximity, etc.). Sound source size is therefore but one of many descriptors that may prove to be useful in understanding felt emotional responses.

V. CONCLUSIONS

The utility of listener-normalized musical affect is largely unknown. In order to develop a complete listener-normalized music perception framework, more research is needed on the perception of psychomechanical aspects of music. By better understanding the perception of sound source attributes within music, more meaningful conclusions may be drawn. In this brief paper, the role of listener-normalized affect has only been discussed in relation to isolated sounds (some of which were musical). Actual music making typically involves constant changes of psychomechanical attributes (energy, proximity, etc.), and therefore, the combination of these changing attributes results in a very complex sound source; perhaps even the most complex sound source regularly encountered in everyday life.

Despite limited knowledge on this topic, we can still speculate about the implications of listener-normalized musical affect, both within and beyond music perception. From a development perspective, listener-normalized emotional responses might be useful for understanding how music perception changes from infancy thru adulthood. To an infant, we should expect that the world "sounds much bigger" than it does to an adult listener. A recent study by Masapollo, Polka, & Menard (2015) found that infants preferred listening to synthesized speech with infant vocal properties relative to adult vocal properties. In this case, listener-normalization (i.e. normalizing sounds to an infant listener) provides an interesting conceptual lens for examining these experimental results. From a theoretical perspective, listener-normalized music affect might be useful for developing a "scaled" theory of emotional responses, such that normalizing psychomechanical musical features to a specific listener might provide insight into how and why that listener prefers certain types of music for certain purposes.

According to Schubert (2013), those most interested in comparing perceived and felt emotional communication tend to be engaged in music perception research. However, the implications for understanding induced emotional responses extend far beyond music into realms such as Human-Computer-Interaction (HCI), comparative psychology, cross-species communication, and potentially even psychological disorders that involve abnormal self-perceptions. It is my hope that researchers investigating emotional responses to music will consider the possibilities afforded by listener-normalization. This concept may be of use not only for future research on induced emotional responses, but also as a meta-analysis tool for investigating variability in previously published work on musical emotions.

ACKNOWLEDGMENT

The author wishes to thank Illinois Wesleyan University for supporting a junior faculty research leave proposal that facilitated this research. The author also wishes to thank Stephen McAdams, Marc Pell, David Huron, and Zachary Silver for their support, feedback, and suggestions during this leave.

REFERENCES

Gabrielsson, A. (2002). Emotion perceived and emotion felt: Same or different? *Musicae Scientiae, 5*(1 suppl), 123-147.

Huron, D., Kinney, D., & Precoda, K. (2006). Influence of pitch height on the perception of submissiveness and threat in musical passages. *Empirical Musicology Review*, *1*(3), 170-177.

Jiang, X., & Pell, M. D. (2015). On how the brain decodes vocal cues about speaker confidence. *Cortex*, *66*, 9-34.

Juslin, P. N. (1997). Perceived emotional expression in synthesized performances of a short melody: Capturing the listener's judgment policy.*Musicae Scientiae*, *1*(2), 225-256.

Juslin, P. N., & Timmers, R. (2010). Expression and communication of emotion in music performance. *Handbook of music and emotion: Theory, research, applications*, 453-489.

McAdams, S., Chaigne, A., & Roussarie, V. (2004). The psychomechanics of simulated sound sources: Material properties of impacted bars. *The Journal of the Acoustical Society of America*, *115*(3), 1306-1320.

Patterson, R. D., Smith, D. R., van Dinther, R., & Walters, T. C. (2008). Size information in the production and perception of communication sounds. In *Auditory perception of sound sources* (pp. 43-75). Springer US.

Masapollo, M., Polka, L., & Ménard, L. (2015). When infants talk, infants listen: pre-babbling infants prefer listening to speech with infant vocal properties. *Developmental Science, 19*(2), 318-328.

Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *American Naturalist*, *111*(981), 855-869.

Schubert, E. (2013). Emotion felt by the listener and expressed by the music: literature review and theoretical perspectives. *Frontiers in Psychology,* 4:837.

Stoelinga, C. (2009). *A psychomechanical study of rolling sounds*. VDM Publishing.

Williams, C. E., & Stevens, K. N. (1972). Emotions and speech: Some acoustical correlates. *The Journal of the Acoustical Society of America*,*52*(4B), 1238-1250.

Emotionally-mediated crossmodal correspondences are sensitive to context

Joshua Peterson[1], Thomas Langlois[1], Stephen E. Palmer[1]
[1]Department of Psychology, University of California, Berkeley, CA, USA

Previous research has provided evidence that music-to-color, music-to-texture, and texture-to-color associations are strongly mediated by shared emotional content (e.g., Palmer, Schloss, Xu & Prado-Leon, 2013; Palmer Langlois & Schloss, 2016). That is, when asked to choose the colors that "go best" with the music they are currently hearing or the textures they are currently seeing, participants tend to choose stimuli from one modality that match a second modality in emotional tone (e.g. *agitated/calm* colors are chosen as going best with *agitated/calm* music and *agitated/calm* textures). In the present experiments, we show that these emotional effects decrease, but are still maintained when using briefer, lower-level musical elements such as timbres and intervals. Emotion mediation is completely absent, however, when using brief, non-musical sounds (e.g., the sound of crinkling plastic or zipping a zipper). Lastly, we show that such differences in emotion mediation are strongly influenced by context, because the same non-musical sounds again exhibit emotion effects when paired with abstract artworks with emotional associations. Combining highly emotional stimuli from two modalities (abstract artworks and full musical pieces) results in the strongest effects of all five experiments.

How and When Sad Music Depends on Empathy

Matthew E. Sachs[1], Jonas Kaplan[1], Morteza Dehghani[1], Antonio Damasio[1], Assal Habibi[1]
[1]Brain and Creativity Institute, University of Southern California, Los Angeles, CA, USA

Sadness experienced in everyday life is largely unpleasant, yet when expressed through music, it is often found pleasurable. Previous research attempting to elucidate this paradoxical concept has shown that liking sad music is correlated with trait empathy, yet which subtype is most predictive of the enjoyment of sad music remains a matter of debate. In addition, how a person's mood, past experiences, and current situations might influence this relationship between empathy and liking sad music has largely been unexplored. To address these questions, we conducted an online survey (N = 429) in which participants reported the type of music they would most likely listen to in various situations and the reasons why they chose that type of music. Multiple regression and structural equation modeling was used to assess the relationship between four subtypes of empathy, emotional responses to sad music, and the enjoyment of sad music in various situations. Results show that fantasy was the strongest predictor of the enjoyment of sad music, and this association was partially mediated by the experience of sublime feelings in response to sad music. Moreover, people who score high on fantasy report liking sad music because it fosters strong, positive feelings, such as feeling touched or moved, and releases negative feelings, such a tension or sadness. Finally, people who prefer listening to sad sounding music when experiencing a negative situation, such as after a breakup or when feeling lonely, score higher on fantasy and personal distress, whereas people who prefer listening to happier sounding music when in such negative situations score higher on perspective taking. The findings allow for a more complete understanding of how trait empathy influences pleasurable responses to negative-valent stimuli within various contexts. Such results could provide new ways of comprehending and treating mood disorders in which the ability to express and experience emotions is attenuated.

Effects of pitch and time expectancy on musical emotion

Sarah Sauvé[1], Aminah Sayed[1], Roger Dean[2] and Marcus Pearce[1]
[1]School of Electronic Engineering and Computer Science, Queen Mary University of London, London, UK
[2]MARCS Institute, University of Western Sydney, Penrith, Australia

Pitch and timing information work hand in hand to create a coherent piece of music; but what happens when this information goes against the norm? While much research in music and emotion explore what emotions are elicited, this work will explore a potential explanation for the elicitation of those very emotions. Meyer (1956) suggested decades ago that musical expectancy triggers emotions in response to music and Juslin and Vastfjall (2008) also propose expectancy as a potential mechanism for musical emotion. The aim of this research is to understand the effect of tonal and temporal structure on expectedness and musical affect as they unfold *in real time*, and whether these effects are independent or not. 40 participants listened to 32 melodies, 16 of which were original and 16 of which were modified. Each group of 16 was categorized by high and low, pitch and time expectancy (as measured by information content). Half the participants continuously rated the level of perceived expectedness on a 7-point Likert scale while the other half rated the perceived arousal/valence on a two-dimensional circumplex model of emotion (Russell, 1980). Participants were further divided in two sub-groups by musical training. A window-based analysis shows that notes with high pitch unpredictability that occur on predictable onset times elicit stronger negative emotions, suggesting that pitch and time information are not processed independently. While expectedness and valence were modulated by the differences in, and manipulations of tonal and temporal structure, arousal was not. Arousal ratings were instead correlated with tempo. Results did not significantly differ between musicians and non-musicians, which can be explained by a shared cultural background and extra-opus knowledge. For a more thorough investigation of the data, time series analysis is underway. Individual differences in rating strategies are also being investigated. Preliminary analysis suggests that tonal and temporal structure do affect perceived expectedness and musical affect as music unfolds over time, and that these structures are processed in a dependent way; however, completion of analysis will solidify the conclusions that can be drawn from this research.

Singing and Moving Increase Social Bonding

K. Harring, L. Edelman, J. Ross, K. Ninivaggi, A. Leeds, K. Mulvey, M. Hogan

***Abstract*— Previous research found that both synchronous movement and music increased perceptions of bonding for targets in a video. Using a 2x3x2 mixed groups design, the current study investigated the effect of having participants engage in music-making and physical movement. Two confederates participated in each condition along with one participant. In half the conditions the participant and the two confederates sang "Row, Row, Row Your Boat" in unison. In the other half of the conditions there was no singing. There were three movement conditions. In one condition, the participant and confederates pumped their arms up and down in unison to the beat of a metronome. In another condition the participant and confederates pumped their arms up and down out of synchrony with each other. In the final movement condition, all three people kept their arms at their sides (no movement). The singing and movement conditions were completely crossed. All conditions were the same duration. At the end of the manipulation participants and confederates completed a questionnaire to measure entitativity, rapport, mood, and manipulation checks. Singing increased bonding compared to not singing. Synchronous movement increased bonding over no movement, but asynchronous movement tended to decrease bonding. Singing alone, moving alone, and singing and moving together all produced similar levels of bonding. Putting both activities together, singing and moving, did not produce greater bonding than either one alone. The asynchronous moving decreased bonding in the singing condition, but was equivalent to no movement in the no singing condition. The results will be discussed in terms of theories of group cohesion, motor neurons, and neurohormonal mechanisms

***Keywords*—** social bonding, synchrony, singing, moving

I. Introduction

Previous research hypothesized that music evolved because of its capacity to promote group cohesion and cooperation [1]. Since music leads to shared emotions, it can facilitates group cohesion, which promotes social bonding [2]. The interaction between group cohesion and cooperation is the foundation for rapport (liking) and entitativity (oneness) in social bonding [3]. Four different characteristics of groups can promote the perception of a single entity. These characteristics are proximity, similarity, common fate and pragnance. If elements are near each other, are similar, move together in the same direction with a similar end point, and form a pattern, respectively, they are more likely to be perceived as one [3]. The more of these characteristics a group demonstrates, the more likely that the group is perceived to be a unit and thus will exhibit a higher level of entitativity. Four different theories can explain how music evolved to support social bonding. The first theory suggests that music causes shared emotion, which encourages social connection [4]. Music can also support synchrony, which also stimulates social connection [5]–[8]. The third theory is that both movement and music trigger mirror neuron responses and those responses stimulate social bonding [9]-[10]. More recently research has suggested that both music and movement trigger neurohormonal mechanisms that stimulate social bonding [11]-[12].

Previous research found that both synchronous movement and music increased perceptions of bonding for targets in a video [13]. In that study participants were simply listening to music and had no active role in making music. Additionally, they watched a video of other people walking synchronously or asynchronously, but again did not actually engage in movement themselves. Even so, they perceived that the synchronous video walkers were socially connected, but the participants did not express a connection to the walkers in the video. [14]. The current study investigated the effect of having participants engage in music-making and physical movement. We assessed the degree to which the participants felt a social bond to the other people in the room who sang or moved with them. There were two singing conditions; either singing or not singing. There were three movement conditions; not moving, moving synchronously, moving asynchronously. We hypothesized that the strongest bonding would be in the condition in which participants both sang and moved in synchrony. We expected that singing alone and moving alone would produce an equal amount of bonding and that the condition with no singing and no moving would not generate perceptions of bonding. Finally we expected that the asynchronous movement would result in the lowest amount of bonding, regardless of the singing.

II. Method

A. Participants

The sample included 91 participants all over the age of 18.

B. Design

A 2x3x2 mixed design was used. The between subjects independent variables were the singing (or not singing) and the moving (not moving, synchronous movement, and asynchronous movement). The within subjects independent variable was the type of rating, entitativity and rapport. The dependent measures were the bonding measures (entitativity and rapport), mood ratings, and manipulations checks.

C. Materials

Participants completed a questionnaire to measure rapport and entitativity. Six questions assessed perceptions of rapport on a series of 7-point Likert-type scales. Examples included, "the people in the room like one another" and "I experience a feeling of togetherness with the people in this room". In addition, four questions assessed perceptions of entitativity on similar scales. Examples included, "to what extent do you think the individuals in the room felt coordinated with each other" and "the people in the room are a unit". Participants indicated the gender with which they identified, how familiar they were with the other participants, and their musical experience. Furthermore, participants were instructed to rate their mood on a

K. Harring is a Dean and Professor of Psychology at Muhlenberg College, L.Edelman is a Professor of Psychology at Muhlenberg College, Allentown PA. (484-664-3426; fax:484-664-????; e-mail: ledelman@muhlenberg.edu).

L. Ross, K. Ninivaggi, A. Leeds, K. Mulvey, and M. Hogan are students at Muhlenberg College, Allentown, PA 18104.

series of 7-point semantic differential scales with varying emotions. These emotions include: happy/sad, negative/positive, excited/calm, confident/insecure, etc. Depending on the condition, manipulation checks evaluated the power of the singing and/or moving manipulations.

D. Procedure

Each condition involved one participant and two experimenters (one male and one female) acting as confederates (synchronous movements refer to the entire group). A third experimenter explained that the experiment was investigating factors that affect attention during studying. Participants were randomly assigned into one of six conditions. The conditions were: (1) singing *Row, Row, Row Your Boat* four times as a group with a metronome and moving one's arms mimicking the arm motions of running in synchrony (singing and synchronous movement); (2) remaining silent with the metronome playing and moving arms as a group (no singing and synchronous movement); (3) singing *Row, Row, Row Your Boat* four times as a group with a metronome while standing still (singing and no movement); (4) remaining silent with the metronome playing while standing still (no singing and no movement); (5) singing '*Row, Row, Row Your Boat*' four times and moving arms whereby the participant was instructed to move in synchrony while confederates purposefully moved out of synchrony with the participant (singing and asynchronous movement); (6) moving arms with a metronome whereby the participant was instructed to move in synchrony with the metronome and the confederates purposefully move out of synchrony with the participant (no singing and asynchronous movement). Participants completed the questionnaire measuring rapport, entitativity, mood, and the manipulation checks. Additionally, they completed the post-experimental questionnaire.

III. RESULTS

The data were analyzed using a three way mixed analysis of variance. Entitativity was rated higher (M= 4.27, SD=1.08) than rapport, M=4.00, SD=1.00, F (1,85)= 6.59, p<.01, η_p^2 =.13. The difference between entitativity and rapport did not interact with any other factor so we considered them together as a measure of social bonding. There was a significant main effect of singing, such that social bonding was greater when participants sang (M= 4.34, SD=0.91) than when they did not sing, M= 3.94, SD=1.13, F(1,85)= 4.62, p=.03, η_p^2 = .05. There was also a significant main effect of moving, such that social bonding was greater when participants moved in synchrony (M= 4.67, SD=0.87) than when there was no movement, (M= 4.07, SD=1.00), and even less when they moved asynchronously, M= 3.68, SD=1.12, F(1,85)= 9.72, p<.01, η_p^2 = .19. LSD post hocs show that the synchronous moving condition resulted in greater bonding than either the no moving or asynchronous moving conditions, p< .01 for both. However, the asynchronous moving condition resulted in only marginally less bonding than the no moving condition, p = .09. There was a significant interaction between singing and moving on ratings of social bonding, F(1,85)= 3.61, p=.03, η_p^2 = .08. Fig.1 shows that singing and moving produced a similar degree of social bonding as singing alone or moving alone. In the control condition with no singing and no moving, social bonding was very low. Moving asynchronously while not singing was similar to the control condition. However, relative to singing with no movement, singing while moving asynchronously greatly decreased the ratings of social bonding.

We conducted several manipulations checks. The mood ratings were averaged together to form a single positive mood scale. There were no significant differences among the conditions for mood. In general, the participants rated their mood as neutral, M= 4.09, SD=0.67.

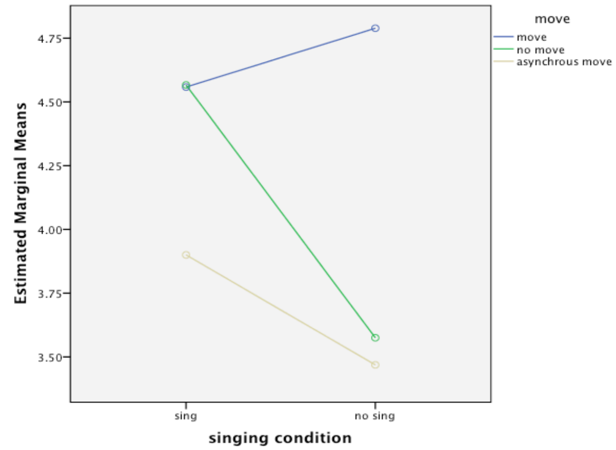


Fig. 1 shows the average rating of social bonding by sining and moving condition.

We also ran correlations between the ratings of rapport and entitativity and the participants' mood, ratings of the song (for those who sang), comfort and embarrassment.. The resulting correlations indicate that for participants who sang, higher entitativity ratings were correlated with higher liking the song, r(45)= .39, p=.01There were no significant correlations between ratings of rapport and any other rating scales.

IV. DISCUSSION

The pattern of results support the prediction that social bonding is low without singing or moving. As predicted singing without moving and moving without singing produced similar results in that both increased social bonding over the control condition. However, combining singing and moving did not increase perceived bonding more than either one alone. Since singing together already involves some synchrony, adding the synchrony of the movement did not have any further effect. More interesting is the result that the asynchronous movement decreased the bonding when singing.

The current results seem to rule out the explanation that singing produces bonding due to a shared positive mood; our participants rated their mood as neutral in all conditions. Social psychological theories of group cohesion based on proximity, similarity, common fate and pragnance could be used to explain these results in that singing together or moving together would result in similarity among members of the group as well as a common fate [3]. This would also help to explain why adding asynchronous movement, which would disrupt similarity and common fate, would decrease social bonding. Unfortunately the mirror neuron hypothesis would work in the exact same way. If hearing music or seeing the movement of others triggers mirror neurons, then copying those actions (singing and moving) should increase the mirror neuron activation. The positive effect on social bonding has been explained as a "shared affective motion experience [15]" or "self-other merging [16]". Conversely, if your movements are out of sync with the movement you see in others, there may be an interference with mirror neuron activity.

Recent research has argued that both music, even passive listening to music, and synchronous movement result in a release of endorphins [15]-[17]. These endorphins, such as oxytocin and vasopressin, have been shown to be related to behaviors that are related to social bonding; increased eye contact [18], generosity [19], and empathy [20]. Thus neurohormonal mechanisms might explain the singing conditions and the synchronous movement conditions in our study, however, they do not account for the results with asynchronous movement. The intensity of the movement and partial synchrony has been studied [21]. High exertion movements without synchrony did not cause a release of endorphins to the same extent as high exertion synchronous movements (as measured by an increase in pain threshold). However, in the same study, low exertion synchronous movements also did not cause an increase in the release of endorphins. In contrast, in other studies such low exertion activities such as synchronized finger tapping resulted in greater social bonding [21].

Given that social coordination and synchrony can promote a sense of belongingness to the group [3], moving out of sync with group members may serve to increase focus on the self, leading to negative self-evaluations, and lower self-esteem [20]. The post-experimental questionnaire revealed that most participants in the asynchronous conditions reported that they were aware that the three individuals were moving their arms at different rates. This awareness may have dampened the need to connect with others who created barriers to coordination with the participant. In other words, the asynchronous movement may have decreased participants' motivation to connect with these peers as a way to restore self-esteem and personal worth.

The current study showed that both singing and synchronous movement led to increased social bonding. Moreover, moving out of sync with other individuals reduces social connection even when the group is singing together. These results appear to imply that asynchronous movement has a more powerful and negative effect on perceptions of groupness; however, we can't make that claim since we did not include a condition where there was asynchronous singing. Future research should further examine the role that both synchrony and asynchrony play in large motor movement and in music-making to more clearly distinguish between the social bonding effects of synchrony alone and of music alone.

IV. References

[1] S. R. Livingstone and W. F. Thompson, "The emergence of music from the Theory of Mind," *Musicae Scientiae*, *Special Issue*, *pp.* 83-115, *2009.*

[2] J. G. Roederer, "The search for a survival value of music," *Music Perception*, vol. 1, no.3, pp. 350-356, 1984.

[3] D. Campbell, "Common fate, similarity, and other indices of the status of aggregates of persons as social entities," *Syst. Res. Behavioral Science*, vol. 3, no. 1, pp. 14-25, 1958. doi:10.1002/bs.3830030103

[4] I. Molnar-Szakacs and K. Overy, "Music and mirror neurons: From motion to 'e'motion," *Social Cognitive And Affective Neuroscience*, vol. 1, no. 3, pp. 235-241, 2006. doi:10.1093/scan/nsl029

[5] D. Lakens and M. Stel, "If they move in sync, they must feel in sync: Movement synchrony leads to attributions of rapport and entitativity," *Social Cognition*, vol. 29, no. 1, pp.1-14, 2011. doi:10.1521/soco.2011.29.1.1

[6] M.J. Hove and J.L Risen," It's all in the timing: Interpersonal synchrony Increases affiliation," *Social Cognition*, vol. 27, no.6, pp. 949-961, 2009.

[7] C.N. Macrae, O. K. Duffy, L. K. Miles, and J. Lawrence, "A case of hand waving: Action synchrony and person perception," Cognition, vol. 109, pp.152-156, 2008.

[8] L.K. Miles, L.K. Nind, and C.N. Macrae, "The rhythm of rapport: Interpersonal synchrony and social perception," *Journal of Experimental Social Psychology*, vol. 45, pp.585-589, 2009.

[9] K. Overy and I. Molnar-Szakacs, "Being together in time: Musical experience and the mirror neuron system," *Music Perception*, vol. 26, no. 5, pp. 489-504, 2009. doi:10.1525/mp.2009.26.5.489

[10] J. R. Matyja, "The next step: Mirror neurons, music, and mechanistic explanation," *Frontiers In Psychology*, vol. 6, 2016.

[11] R.M. Dunbar, K. Kaskatis, I. MacDonald, and V. Barra, "Performance of music elevates pain threshold and positive affect: Implications for the evolutionary function of music," *Evolutionary Psychology*, vol. *10*, no. 4, pp. 688-702, 2012.

[12] L. Gebauer, M.L. Kringelbach, and P. Vuust, "Ever-changing cycles of musical pleasure: The role of dopamine and anticipation," *Psychomusicology: Music, Mind, and Brain*, vol. 22, no. 2, pp. 152-167, 2012.

[13] L.L. Edelman and K.E Harring, "Music and social bonding: The role of non-diegetic music and synchrony on perceptions of videotaped walkers," *Current Psychology: A Journal For Diverse Perspectives On Diverse Psychological Issues*, 2014. doi:10.1007/s12144-014-9273-y

[14] L.L. Edelman, K. E. Harring, "Music bonds the video walkers but not the participants," Unpublished manuscript, 2014.

[15] B. Tarr, J. Launay, and R. M. Dunbar, "Music and social bonding: 'Self-other' merging and neurohormonal mechanisms," *Frontiers In Psychology*, vol.5*, 2014.*

[16] U. Nilsson, "Soothing music can increase oxytocin levels during bed rest after open-heart surgery: A randomised control trial," *Journal Of Clinical Nursing*, vol. 18, no. 15, pp. 2153-2161, 2009. doi:10.1111/j.1365-2702.2008.02718.x

[17] D. Huron, "Is music an evolutionary adaptation?" in R. J. Zatorre, I. Peretz, R. J. Zatorre,and I. Peretz (Eds.) , *The biological foundations of music*, New York, NY, US: New York Academy of Sciences, pp. 43-61, 2001.

[18] K. MacDonald and T. M. MacDonald, "The peptide that binds: A systematic review of oxycotin and its prosocial effects in humans," *Harvard Review Of Psychiatry*, vol. 18, no.1, pp. 1-21, 2010. doi:10.3109/10673220903523615

[19] J.A. Barraza and P.J. Zak, "Empathy toward strangers triggers oxytocin release and subsequent generosity," in O. Vilarroya, S. Altran, A. Navarro, K. Ochsner, A. Tobeña, O. Vilarroya, ... A. Tobeña (Eds.) , *Values, empathy, and fairness across social barriers* , New York, NY, US: New York Academy of Sciences, pp. 182-189, 2009.

[20] J. Lumsden, L.K. Miles, and C.N. Maccrae, "Sync or sink? Interpersonal synchrony impacts self-esteem," *Frontiers In Psychology*, vol. 5*,* 2014.

[21] Oullier O, de Guzman G, Jantzen K, Lagarde J, Kelso J. Social coordination dynamics: Measuring human bonding. *Social Neuroscience* [serial online]. 2008;3(2):178-192.

The Expression of Anger, Angst, Violence and Power in Music: Aesthetic, Sociological, and Evolutionary Perspectives

Nathaniel Condit-Schultz[1]
[1]School of Music, Ohio State University, Columbus, Ohio, USA

Of all music genres, heavy metal and gangsta rap are the most socially, politically, and aesthetically polarizing, with a deep divide between their dedicated fan-bases and the larger community which cannot comprehend the appeal of music which seems to be uniformly angry, vulgar, and violent. This paper will present a theory of emotion and aesthetic in these musical genres which is grounded in evolutionary psychology. I will argue that the core aesthetic appeal of these genres is not anger or violence but the expression of a feeling of being *powerful,* with deep evolutionary roots in male-male competition. Sexual competition is an important driver of evolutionary selection and has thus shaped human nature and culture. As a result, the psychology of human males, especially adolescents, is heavily influenced by an innate desire to dominate—to feel physically, mentally, socially, and sexually powerful. I will discuss the artistic and aesthetic roles of these feelings in music, as well as their ethical and social ramifications.

I propose that the prized aesthetic of *heaviness* in metal can be understood as an expression of power. Likewise, the *gangsta* image cultivated by rappers is a textbook depiction of masculine power and sexual dominance. Rappers express physical power through violent imagery and boasting. Metal musicians instead express physical power through raw energy and acoustic cues which are associated with strength and aggression. Metal and rap also express competitive domination through their emphasis on mental acuity and skill: The virtuoso rock guitarist is the poster child of sexual selection in the music. It has been demonstrated that men unconsciously show off their vocabulary when courting women—a precursor to rap's skillful word-play. Studies have also observed that courting men increase their displays of non-conforming behavior and risk taking, both of which are expressed in these genres' rebellious attitudes and rejections of authority.

A new measure of the well-being benefits of music participation

Amanda E Krause[1], Jane W Davidson[2], and Adrian C North[1]
[1] School of Psychology and Speech Pathology, Curtin University, Perth, WA, Australia
[2] ARC Centre of Excellence for the History of Emotions, Faculty of VCA & MCM, the University of Melbourne, Melbourne, VIC, Australia

A surge of interest in the impact of music on well-being leaves researchers to consider the benefits of music participation. Although prior qualitative studies have begun to outline the benefits, these studies are limited by narrowed foci and small samples. Additionally, research that considers both vocal and instrumental participation is scant. The present study developed a new measure to scrutinize the potential well-being benefits from musical participation via (1) an extensive literature review and qualitative meta-analysis and (2) scale refinement and testing. A literature review generated a total of 1983 well-being items, which was reduced to a list of 562 after eliminating redundant and analogous items. Then four judges rated the suitability of the items, and, in turn, 40 distinct items were retained for scale construction. Participants were asked to "consider the benefits to your well-being that result from your participation in music" and rate accordingly how strongly they agreed with each item on a 7-point Likert scale. The underlying structure of the measure was identified via a Principal Axis Factor analysis with Promax rotation (N=207). Analysis uncovered five dimensions: "mood and coping" (concerning feeling uplifted, feeling better, and relieving stress), "esteem and worth" (concerning feeling positive about life, a sense of belonging, and an increase in self-esteem), "socializing" (related to bonding with other people and connecting with a community), "cognitive" (about thinking and processing), and "self-actualization" (reflecting self-improvement). Extending previous findings, this new measure offers a comprehensive tool for researchers investigating music participation. Additional analyses (underway) which consider the measure's construct validity relative to the MUSE and criterion validity relative to the Satisfaction With Life Scale will also be discussed in the presentation.

Musical Intensity in Affect Regulation: Interventions in Self-Harming Behavior

Diana Hereld[1]
[1]Music, University of California, San Diego, CA, USA

Prior research associates listening to heavy music with reduced suicide risk, especially among teenage girls when utilized for vicarious release. Nevertheless, few studies consider the active use of heavy music in self-regulation for those who suffer from thoughts of self-harm and/or mental illness. In order to to better understand the mechanisms by which engaging with heavy and intense music may circumvent self-harming behavior, a pilot study is presented of 283 subjects. The majority of those surveyed report suffering from thoughts of self-harm or mental disorders. To examine the use of affect regulation via both generic (non-specified) and heavy, intense, and highly emotive music, we created the Music in Affect Regulation Questionnaire (MARQ), utilizing music in mood regulation (MMR) strategies from the work of Saarikallio. We identify heavy music by the presence of capacious, distorted riffs; loud, pervasive percussion; or an overall feeling of 'raw power,' emotion, and affective intensity stemming from the instrumental or vocal parts. Our findings collectively show that heavy music listeners (and those who have thoughts of self-harm, in particular) interact with definitively heavy, intense, or highly emotive music differently than with generic music, especially in the use of modulating negative mood. These findings seem less related to genre-specific categories than certain musical commonalities collectively understood as intensity, and provide significant evidence for heavy music's ability to circumvent self-destructive impulses, especially when applied in tandem with specific listening strategies of affect-regulation. Additional evidence from prior case studies further suggests the value of deeper investigation of the conscientious use of heavy music as a potential intervention for those suffering from affect dysregulation and self-harm.

ISBN 1-876346-65-5 © ICMPC14

Music Engagement, Personality, Well-Being and Affect in a Sample of North Indian Young Adults

A. Chakraborty, D. K. Upadhyay, R. Shukla

Abstract—The current research, comprised of two studies, explored the nature of music engagement and its relationship with various psychological constructs in a sample of North Indian young adults. **Study 1** examined the relationship between music preferences (MP), listening styles (LS), functions of music (FoM), perceived *rasa* (music), and personality traits (PT - Big Five Factors). A sample of 77 young adults (M = 39; F = 38; M = 22.7 years) completed measures of the above constructs and data were analysed via correlations, one-way ANOVA, post hoc tests, and T-tests. Significant correlations were found between LS & PT; MP & PT; FoM & PT; FoM & emotion; and LS & FoM. Findings indicated stronger preferences for genres namely Romantic songs, Soft songs, and Filmy (Sad) songs. Gender difference existed in terms of MP, perceived emotions and LS. Music listening mainly served as a *'source of pleasure and enjoyment'* and which *'calms, motivates, or reminds of past events'*. Musical genres inducing *santoṣa rasa* were perceived significantly higher in female participants. Based on the findings, a 'Music Engagement Model (MEM) for Young Adults' describing their music behaviour, have been proposed.

Study 2 examined affect (thoughts, feelings and actions) while listening to the lyrics of various music genres in young adults with varied personality traits. This study further examined which personality factor was more associated with thoughts, feelings and action tendencies generated through songs of various genres. A sample of 60 young adults (30 boys and 30 girls) of age group 18-27 years, pursuing graduation and post-graduation degrees from Amity University, Lucknow campus were selected. Tools used were Big Five Inventory (by John & Srivastava) and a 4-point Cognitive, Affective and Conative (CAC) scale (developed by Authors). Findings provide insights about the significance of music as media in day-to-day lives of young adults, particularly on their cognition and the amount of affect based on their personality factors. The time phrase i.e. young adulthood, what has been called the most crucial age, needs to be exposed to such music which does not only prove to be a source for chills and enjoyment but also which fosters its well-being.

***Keywords*—**Music preference; *rasa*; personality; psychological well-being; affect.

Ms A. Chakraborty, student of MA (Clinical Psychology), is with the Amity Institute of Behaviour and Allied Sciences, Amity Uttar Pradesh, Lucknow, India, 226028 (corresponding author to provide phone: +91-7071758062; e-mail: ahelic28@gmail.com).

Dr. D. K. Upadhyay, Assistant Professor, is with the Amity Institute of Behaviour and Allied Sciences, Amity Uttar Pradesh, Lucknow, India, 226028 (e-mail: dkupadhyay@lko.amity.edu).

Dr. R. Shukla, Assistant Professor, is with the Amity Institute of Behaviour and Allied Sciences, Amity Uttar Pradesh, Lucknow, India, 226028 (e-mail: rshukla1@lko.amity.edu).

I. Introduction

All humans across all cultures are exposed to music and potentially possess the innate ability to understand and respond to music. In this modern era, music is so pervasive it is unavoidable. The music of India (one of the oldest unbroken musical traditions in the world), is inextricably interwoven not only with ritualistic and devotional side of religious lives but also with day-to-day life experiences. Music accompanies a person from birth until death. However, we have little understanding of how people use and experience music in their daily lives [3], and how cultural background, gender, personality, listening habits and other factors may influence the uses of music. Involvement in musical activities has been shown to have positive effects on mood [22], quality of life [4] and engagement [5], and to be a very rewarding leisure activity [12]. A challenge to such an investigation is that music is used for many different purposes.

In recent years, researchers have been particularly interested in adolescents and young adults who are marginalized and/or experiencing major psychological issues and have found that they prefer heavier forms of music such as heavy metal and hard rock [21] [6] [23]. It is presumed that these music preferences reflect their values, conflicts, and developmental issues with which these youth are dealing. Music listening is one of the most enjoyable activities reported by young adults.

Karl Mannheim relates young adulthood with the ability to question and reflect upon life and experiences. To him, this time begins about the age of seventeen [13]. He justifies his age differentiation in further expressing the importance of living in the "present," "the up-to-dateness of youth therefore consists in their being closer to the 'present' problems" (pp. 300–301). "The time phrase [young adulthood] embraces what has been called the most crucial age range for the creation of a distinctive and self-conscious generation" [8]. Young adults are those who have reached sexual maturity, but are not married [18].

Young adults' subsequent transition to adulthood is notoriously stressful, placing demands on young people's coping resources and putting them at greater risk of developing mental health problems [1] [11]. There is, therefore, a need to investigate everyday strategies that young people use to support their well-being. There is a growing recognition amongst scientists that musical behaviour is central to our humanity, to what it means to be human. Not least, we engage in musical activities because doing so sweetens and structures our leisure time and thereby makes us happy and increases our well-being [7] [15]. All these assumptions, however, lack an explanation for *how* music can lead to such positive effects [17].

Viewed against this backdrop, the current research, comprised of two studies, explored the nature of music engagement and its relationship with various psychological constructs in a sample of North Indian young adults. **Study 1** examined the relationship between music preferences (MP), listening styles (LS), functions of music (FoM), perceived *rasa* (emotion), personality traits (PT - Big Five Factors) and psychological well-being (PWB). **Study 2** examined the impact of various songs (with lyrics) of various genres had any effect on the thoughts (cognitive), feelings (affective) and actions (conative) of young adults with varied personality traits.

II. Study I

A. Method

Participants

A total of 77 respondents (male = 39; female = 38 with mean age = 22.7 years), pursuing UG and PG programs from different schools/institutions of Amity University, completed the survey.

Measures

Demographic Data Sheet: The data sheet was used to gain demographic details of the respondents (e.g. age, gender, socioeconomic status, and music background).

Music Preference Scale: Music Preference Scale developed by the researcher was used to figure out the music preference of the respondents. The scale included 23 music genres, to be rated on a seven-point Likert rating scale by the respondents to indicate their preference for listening to a particular music genre. This scale also included one open ended item asking respondents to add, if any, to the already listed genres. Each music genre was accompanied by one open ended item asking them to respond to *'when (time, place, mood, etc.) do you prefer listening to this music genre?'* Cronbach's alpha of the scale has been found to be 0.85.

Music Engagement Scale: This Likert type rating scale was used to assess the music engagement of the participants [24]. It covered items pertaining to number of hours devoted per week by each participant and their preferred ways of listening to music. *Functions of Music Scale:* This Likert type rating scale [19] comprised of 10 items was used. Cronbach's alpha for this scale has been 0.76.

Music Emotion Scale: This rating scale, proposed by author, has been adopted from Koduri and Indurkhya (2010) [10] and Misra (2014) [14]. It has 11 *rasa* (with clusters of emotions) to be rated on seven point scale. The reliability coefficient (Cronbach's alpha) is 0.83.

Big Five Inventory (BFI): To ascertain the personality type of the respondents 'Big Five Inventory' developed by John and Srivastava (1999) [9] was used. The BFI contains five subscales: extraversion, agreeableness, conscientiousness, neuroticism, and openness. The inventory seeks responses on a 5-point rating scale ranging from ''totally agree" to "totally disagree". Certain items in the inventory are reverse scored. Reliability of the BFI ranges from 0.79 to 0.88.

Survey

Based on previous literature, for example on musical engagement [24], *rasa* [10] [14], personality factors [9], and music preferences (Upadhyay, in press), a survey with ratings scales was administered individually on the participants.

B. Results and Discussion

Results revealed that majority of 55.90% respondents actively listened to music up to 6 hours, 27.3% listened between 6 to 10 hours, and 13% listened more than 10 hours per week. Active listening was defined as 'listening carefully or on purpose, instead of doing other things as well'. Only 3.90% respondents did not actively listen to music. Passive listening was defined as 'listening in the background, while you do other things'. 53.3% respondents listened up to 6 hours, 19.5% between 6 to 10 hours, and 24.5% listened more than 10 hours per week in the background. 2.6% indicated 'not (listening) at all' in the background. The male respondents ($M = 4.33$, $SD = 1.91$), in comparison to female respondents ($M = 3.55$, $SD = 1.69$) were more engaged in active listening.

Respondents' responses to the open-ended question *'when do you prefer listening to particular music genre'* were put in three categories: *time*, *context* and *mood*. Respondents indicated that they listened to Bollywood sad (slow) songs anytime (50%) being alone (100%) and when they were in bad (low) mood (60%) or felt gloomy/sad (40%). On the other hand, most of the respondents (87.5%) enjoyed listening to romantic (love) songs being alone but when they were in happy mood (90%). Genres namely rock songs (66.67%), hip-hop (85.71%), remix (57.14%) and rap (57.14%) were preferred in clubs/parties. Bhajan (88.89%) was preferred in morning whereas patriotic songs (100%) on special occasions. Respondents also showed preferences for particular genres in terms of actions they are involved in while listening to them. While *driving* respondents generally preferred romantic (love) songs, melodious film songs, English songs, sufi and rock songs; when felt like *dancing*, indicated for rap, hip-hop, pop, remix, new age, Punjabi and rock songs; and to *relax* they preferred listening to genres like trance, instrumental, sufi, classic, jazz and *ghazal*.

Findings indicated preference for listening to selected music genres. Results of a one-sample test against the centre of the 7-points rating scale (test value = 4) indicated whether ratings were higher or lower from 'somewhat enjoyable'. Romantic/love songs as well as film songs were generally favoured over other genres. On seven-point scales, mean ratings of 'romantic/love songs' were significantly higher in female respondents ($M = 6.00$, $SD = 1.21$) than male respondents ($M = 5.33$, $SD = 1.61$). Whereas, mean ratings of 'rock song', 'pop', 'blues', 'trance', 'jazz', and 'instrumental music' were significantly higher in male respondents than female respondents.

Table of Contents for this manuscript

Respondents also indicated their listening styles on seven-point scale. The highest ratings were yielded for ‘emotional listening’ (i.e., experiencing personal emotional responses to music, M = 4.74, SD = 1.85), followed by ‘moving the body or parts of the body’ (M = 4.31, SD = 1.88), ‘having visual or other associations’ (M = 4.25, SD = 1.79), and ‘analytical listening/concentrating on the structure’ (M = 3.94, SD = 1.89). Moreover, mean ratings of ‘emotional listening’ were significantly higher in female respondents (M = 5.26, SD = 1.65) than male respondents (M = 4.23, SD = 1.99), t (77) = 2.47. Mean ratings of ‘moving the body or parts of the body’ were also higher in female respondents (M = 4.87, SD = 1.60) than male respondents (M = 3.77, SD = 1.99), t (77) = 2.67.

Music genres were correlated with personality factors. Genres namely rock song, ghazal, patriotic songs, sufi, classic, hip-hop, English songs, blues, jazz, trance and instrumental music were significantly correlated with ‘openness’. ‘Extraversion’ was correlated with genres namely patriotic songs, melodious film songs and folk; whereas ‘neuroticism’ was negatively correlated with rock and rap music genres and ‘conscientiousness’ with English songs.

Significant correlations with listening styles were found for two personality factors. Neuroticism’ was correlated with ‘emotional listening’, r (77) = .28, $p < .05$, and ‘openness’ with ‘analytical listening’, r (77) = .27, $p < .05$.

Respondents were asked what functions active or passive music listening have for them. Results indicated that music mainly serves as a ‘source of pleasure and enjoyment’, ‘calms/releases stress/relaxes’, ‘motivates’ or ‘reminds of past events’. Significant correlations were found between hours spend in active listening and music that ‘motivates’, r (77) = .25, $p < .05$, and between passive listening and music that ‘reminds of past events’, r (77) = .38, $p < .01$.

Respondents rated their experienced *rasa* while listening to music. The highest ratings were yielded for *Śṛṅgāra rasa* (represents emotions like arousal, longing, desire, naughty, romance and love, M = 5.16, SD = 1.79), followed by *Santoṣa rasa* (represents emotions like joy, pleasure, excitement, and contentment, M = 4.58, SD = 2.10), and *Śānta rasa* (represents emotions like steady, rest, and peace, M = 4.25, SD = 2.07). Significant difference in rating of *Santoṣa rasa* were found between female respondents (M = 4.05, SD = 2.43) and male respondents (M = 5.10, SD = 1.59), t (77) = 2.25, $p < .05$.

Listening style ‘visual and other associations’ was found significantly correlated with five functions served by music listening, which were as follows: *It reminds me of past events,* r (77) = .33, $p < .01$; *It evokes visual images,* r (77) = .42, $p < .01$; *It moves to tears/chills/other bodily reactions,* r (77) = .24, $p < .05$; *It functions as catharsis,* r (77) = .44, $p < .01$; and *It motivates,* r (77) = .34, $p < .01$. This listening style was also found significantly correlated with four *rasa* namely *Śṛṅgāra* (r = .34, $p < .01$), *Bhayānaka* (r = .25, $p < .05$), *Adbhuta* (r = .27, $p < .05$), and *Santoṣa* (r = .25, $p < .05$) evoked by music listening. It indicates that respondents of this listening style could be more engaged and affected by the music they listen to.

Significant correlations with ‘openness’ personality factor were found with five functions served by music listening.

Previous researches argue that after exercise, music is the second most commonly used mood-regulation strategy in young people [20]. Various forms of musical engagement were related to well-being through emotion regulation strategies [16]. Although, this study didn’t investigate correlations between young adults’ musical engagement and their psychological well-being. However, it can be hypothesized that young adults’ day-to-day musical engagements may predict their psychological well-being.

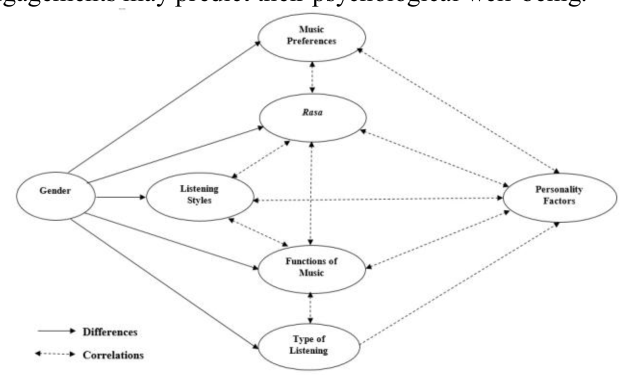


Figure 1: Music Engagement Model (MEM) for Young Adults

This study proposes a tentative **Music Engagement Model (MEM) for Young Adults** (see Figure 1), which indicates gender differences in terms of music preferences, listening styles, types of listening, functions of music and *rasa* perceived. Music preferences and listening styles are correlated with the *rasa* perceived. Functions served by music listening are correlated with type of listening, listening styles and the *rasa* perceived. Personality factors are correlated with the *rasa* perceived, music preferences, functions of music, listening styles and type of listening (active/passive). Moreover, music listening occur in a specific psychosocial context (e.g., place, time, mood etc.). Proposed model assumes that variables like types of *rasa*, preferences for music genres, listening styles and functions served by music listening may predict psychological well-being. The proposed model and its assumption are subject to verification.

C. Conclusion

In conclusion, investigating young adults’ musical engagement, their background, and their personality enhances our understanding of their musical behaviour. This study puts forth several questions to be answered. For example, how do they develop preference for music genres? How do they use music strategically to regulate their moods/emotions or to deal with their sufferings? When do they prefer listening to a particular music? How many hours they spend listening to particular music genre in a day or in a week? Future research could explore the musical experiences of young adults of both music and non-music background through in-depth interviews. Answers to these questions may help researchers to explain the musical behaviour of young adults that support their well-being.

III. Study II

A. Methods

Participants

60 participants (M = 30; F = 30) of the age 18-27 years (M=20.42yrs SD=2.12) pursuing various degrees at both graduation and post-graduation levels from various parts of North India were enrolled by using a snowball-sampling technique.

Measures

Big Five Inventory (BFI) – O.P. John and S. Srivastava (1999) constructed this 44 item inventory with 0.75 reliability that measures an individual on the Big Five Factors (dimensions) of personality, namely, Openness to Experience; Conscientiousness; Extraversion; Agreeableness and Neuroticism. (Goldberg, 1993).

Cognitive, Conative and Affective (CCA) Scale – The researcher constructed this 4-pointer (in order to avoid neutral responses) rating scale which measured the perceived thoughts, feelings and actions while listening to each music piece.

Procedure

Firstly, 60 participants of the final study had to fill in their socio-demographic details and the BFI. They were then presented with chits, which contained the serial number of a specific genre. The various genres were numbered in the following way: 1= Item Songs; 2= Romantic genre; 3= Rap or Outrageous; 4= Sufi Genre; 5= Value-based Songs. Set Randomization method within the genres was used to nullify the practice effect and according to the order of the chits, songs for each genre were presented to them. Each song was played for a minute and they were asked to rate the amount of thoughts, feelings and actions that were evoked on listening to that particular piece according to the 4-pointer rating scale where, 1= Not at All; 2= somewhat; 3= Much; 4= Very Much. Each participant had to rate all the 20 songs and they could leave their feedback if they wished to at the end of the survey.

B. Results and Discussion

On an average, 25.44% respondents were found above 75th percentile who rated high on all three categories namely thought, feelings, and action after listening to songs from various music genres. Particularly for the genre 'Item', 31.67% respondents rated the highest for feelings and only 13.33% respondents rated the least for thoughts.

Ratings for thoughts evoked by genre 'Romantic' was the highest (M=74.56, SD=18.57) which is followed by 'Rap', 'Item' and 'Sufi' genres. Feelings were evoked maximally (M=65.03, SD=12.97) by the genre 'Romantic' and then by 'Sufi' genre (M=58.01, SD=10.23). As far as actions are concerned, genres namely 'Romantic', 'Rap/Outrageous', and 'Item' were respectively found more action evoking.

Results indicated significant gender difference in terms of thoughts {$t(58) = -2.545$, $p < .014$} evoked by Rap/Outrageous songs and feelings {$t(58) = 2.908$, $p < .005$) evoked by Value based songs. Romantic songs evoked more amount of thoughts (M=79.03, SD=16.64), feelings (M=67.33, SD=12.11) and actions (M=64.83, SD=13.44); whereas, Rap/Outrageous as well as Sufi songs evoked more amount of thoughts (M=70.10, SD=17.32; M=66.03, SD=13.22) in girls respectively. However, Value based songs produced more feelings (M=56.63, SD=12.81) in boys as compared to girls.

However, for genre Romantic, mean differences are significant only for thoughts {$t(58) = -2.55$, $p < .013$} and actions {$t(58) = -2.69$, $p < .009$}. Moreover, across all music genres, except thoughts produced by Value based songs, it is apparent that Intermediate passed students perceived more thoughts, feelings and actions than UG/PG students after listening to songs of different genres.

Thoughts evoked by Items songs were positively correlated with 'extraversion' personality factor, $r(60) = .298$, $p < .021$. Thoughts ($r(60) = .282$, $p < .029$) and feelings ($r(60) = .341$, $p < .008$) perceived while listening to Romantic songs were positively correlated with 'neuroticism', whereas, thoughts ($r(60) = -.302$, $p < .019$) and feelings ($r(60) = -.289$, $p < .025$) produced from Romantic songs were negatively correlated with 'conscientiousness.'

'Openness' and 'extraversion' were positively correlated with thoughts {$r(60) = .328$, $p < .011$, $r(60) = .415$, $p < .001$} and feelings {$r(60) = .289$, $p < .025$, $r(60) = .362$, $p < .004$} rated for genre 'Rap/Outrageous' respectively, while, 'conscientiousness' was negative correlated with thoughts, $r(60) = -.289$, $p < .025$, and actions, $r(60) = -.288$, $p < .026$.

Feelings, $r(60) = -.361$, $p < .005$, evoked by Sufi songs were negatively correlated with 'conscientiousness' whereas thoughts, $r(60) = .335$, $p < .009$, were positively correlated with 'neuroticism.' As far as Value based songs were concerned, feelings, $r(60) = -.276$, $p < .033$, generated were negatively correlated with 'agreeableness.'

C. Conclusion

This study provides insights about the significance of music as media in day-to-day lives of young adults, particularly on their cognition and the amount of affect based on their personality factors. The time phrase [young adulthood] embraces what has been called the most crucial age which range for the creation of a distinctive and self-conscious generation [8], needs to be exposed to such music which not only proves to be a source of chills and enjoyment but also which may foster well-being.

Due to paucity of time, the sample size was restricted to only 60 participants only. The paper was also limited to only finding out the impact of various music genres on the thoughts, feelings and actions of the young adults. The dominant thought, feeling or action for each genre could not be derived. The research was done using only five genres out of various music genres due to time constraints. Including many other genres might throw light on a different dimension altogether. Whether previous music training had any significant contribution in evoking thoughts, feelings and actions also could not be studied.

REFERENCES

[1] J. J. Arnett, "Adolescents' uses of media for self-socialisation," *J. You. and Ado.,* vol. 24, pp. 519–533, 1995.

[2] J. J. Arnett, *Adolescence and emerging adulthood: A cultural approach,* 4th Ed. Boston: Prentice Hall, 2010.

[3] T. Chamorro-Premuzic, and A. Furnham, "Personality and music: Can traits explain how people use music in everyday life," *British Journal of Psychology.,* vol. 98, pp. 175–185, 2007.

[4] S. Clift, G. Hancox, I. Morrison, B. Hess, G. Kreutz, and D. Stewart, "Choral singing and psychological wellbeing: Quantitative and qualitative findings from English choirs in a cross-national survey," *Journal of Applied Arts and Health,* vol. 1, no. 1, 2010, pp. 19–34

[5] J. W. Davidson, "Musical participation: Expectations, experiences, and outcomes," in *Music and the Mind: Essays in Honour of John Sloboda,* I. Deliege and J. W. Davidson, Eds. Oxford, UK: Oxford University Press, 2011, pp. 65–87.

[6] C. H. Hansen, and R. D. Hansen, "Schematic information processing of heavy metal lyrics," *Communication Research.,* vol. 18, pp. 373 – 411, 1991.

[7] P. Hills, and M. Argyle, "Musical and religious experiences and their relationship to happiness," *Personality and Individual Differences.,* vol. 25, pp. 91-102, 1998.

[8] K.M. Jennings, and R. G. Niemi, *Generations and Politics.* New Jersey: Princeton University Press, 1981.

[9] O. P. John, and S. Srivastava, "The big-five trait taxonomy: History, measurement, and theoretical perspectives," in *Handbook of Personality: Theory and Research,* 2nd ed. L.A. Pervin & O.P. John Eds. New York: Guilford Press, 1999.

[10] G. P. Koduri, and B. Indrukhya, "A behavioural study of emotions in South Indian classical music and its implications in music recommendation systems," in *SAPMIA '10 - Proceedings of the 2010 ACM workshop on Social, adaptive and personalized multimedia interaction and access,* 2010, pp. 55-60.

[11] S. Laiho, "The psychological functions of music in adolescence," *Nordic Journal of Music Therapy,* vol. 13, pp. 47–63, 2004.

[12] A. Lamont, "University students' strong experiences of music: Pleasure, engagement and meaning," *Musicae Scientiae,* vol. 15, no. 2, pp. 229–249, 2011.

[13] K. Mannheim, "The problem of generations," in *Essays on the sociology of knowledge*, P. Kecskemeti, Ed. London: Routledge and Kegan Paul, 1952, pp. 276-320.

[14] G. Misra, "Psychology of emotions: Some cultural perspectives," in *Foundations and applications of Indian psychology*, R. M. Matthijs Cornelissen, G. Misra and S. Varma, Eds. New Delhi: Pearson, 2014, pp. 205-222.

[15] S. Pinker, *How the Mind Works*. New York: Norton, 1997.

[16] N. S. Rickard, "Music listening and emotional wellbeing," in *Lifelong engagement with music: Benefits for mental health and wellbeing,* N. S. Rickard and K. McFerran, Eds. New York: Nova Science Publishers Inc., 2012, pp. 207- 238.

[17] T. Schafer, and P. Sedlmeier, "What makes us like music? Determinants of music preference," *Psychology of Aesthetics, Creativity, and the Arts*., vol. 4, no. 4, pp. 223-234, 2010.

[18] N. Schindler, "Guardians of disorder: Rituals of youthful culture at the Dawn of the modern age," in *Ancient and medieval rites of passage: A history of young people in the West,* G. Levi and J. C. Schmitt, Eds. London: The Belknap Press of Harvard University Press, 1997, pp. 259–273.

[19] J. A. Sloboda, "Everyday uses of music: a preliminary study," *Music, Mind, & Science,* S. W. Yi, Ed. Seoul: Seoul National University Press, 1999/2005, pp. 354-369. Reprinted in Sloboda, J. A. *Exploring the music mind: Cognition, emotion, ability, function*. Oxford: Oxford University Press, 2005, pp. 319-331.

[20] R. E. Thayer, R. Newman, and T.M. McClain, "Self-regulation of mood: strategies for changing a bad mood, raising energy, and reducing tension," *Journal of Personality and Social Psychology.,* vol. 67, pp. 910-925, 1994.

[21] K. J. Took, and D.S. Weiss, "The relationship between heavy metal and rap music on adolescent turmoil: Real or artefact?" *Ado.,* vol. 29, no. 115, pp. 613-621, 1994.

[22] E. Valentine, and C. Evans, "The effects of solo singing, choral singing and swimming on mood and physiological indices," *British Journal of Medical Psychology.,* vol. 74, pp. 115–120, 2001.

[23] H. Wass, M. D. Miller, and R. G. Stevenson, "Factors affecting adolescents' behaviour and attitudes towards destructive rock lyrics," *Death Studies.*, vol. 13, pp. 287 – 303, 1989.

[24] C. Wöllner, C, J. Ginsborg, and A. Williamon, "Music researchers' musical engagement," *Psychology of Music.,* vol. *39*, no. 3, pp. 364-382, 2011.

Music rehearsals and well-being: what role plays personality?

Eva Matlschweiger[1], Sabrina Sattmann[1] and Richard Parncutt[1]
[1]Center for Systematic Musicology, University of Graz, Graz, Austria

Numerous studies have shown that personality traits play an important role in determining well-being, especially extraversion being a strong predictor of positive affect (e.g. Steel, Schmidt & Shultz, 2008). In music, positive effects of (choral) singing on well-being have been reported (e.g. Beck et al., 2000). We explored the relationship between personality and well-being in three amateur choirs (n = 57, 44 female, mean age 59.7 years), two amateur brass bands (n = 54, 20 female, mean age 34.1 years) and three amateur theater groups (n = 34, 21 female, mean age 32.1 years). Participants completed the Positive Negative Affect Schedule (PANAS), the Perceived Stress Questionnaire (PSQ) and the State-Trait Anxiety Inventory (STAI; state questionnaire only) before and after a 1.5-hour rehearsal, as well as the short version of the Big Five Inventory (BFI-K) after the session. We found differences between groups for extraversion and openness to experience, indicating that members of the theater groups are more extraverted than choir singers and that members of the brass bands are less open to experience than those of choirs and theater groups. Multiple linear regression analyses show that the difference score for positive affect between pre- and post-measurements can be predicted by extraversion; a higher score on the extraversion scale accounts for a bigger increase of positive affect. Members of the theater groups benefit more than brass band members. As for the difference score of STAI, higher ratings for openness to experience were found to predict the decrease of anxiety. Anxiety is predicted to decrease more for members of the theater groups than members of the brass bands. Results are of relevance for hospitals, retirement homes, and therapists that organize or recommend active musical activities for their clients.

Motivating Stroke Rehabilitation Through Music: A Feasibility Study Using Digital Musical Instruments In The Home

Pedro Kirk[1,2], Mick Grierson[1], Rebeka Bodak[2,3,4], Nick Ward[5], Fran Brander[5], Kate Kelly[5], Nickolas Newman[6], Lauren Stewart[2,3,4]
[1]Department of Computing, Goldsmiths University of London, London, UK
[2]Department of Psychology, Goldsmiths University of London, London, UK
[3]Center for Music in the Brain, Dept. of Clinical Medicine, Aarhus University, Denmark
[4]The Royal Academy of Music, Aarhus/Aalborg, Denmark
[5]The National Hospital for Neurology and Neurosurgery, Queen Square, London, UK
[6]Hounslow and Richmond Community Health Care NHS Trust, London, UK

Digital approaches to physical rehabilitation are becoming increasingly common and embedding these new technologies within a musical framework may be particularly motivating. The current feasibility study aimed to test if digital musical instruments (DMIs) could aid in the self-management of stroke rehabilitation in the home, focusing on seated forward reach movements of the upper limb. Participants (n=3), all at least 11 months post stroke, participated in 15 researcher-led music making sessions over a 5 week intervention period. The sessions involved them 'drumming' to the beat of self-chosen tunes using bespoke digital drum pads that were synced wirelessly to an iPad App and triggered percussion sounds as feedback. They were encouraged to continue these exercises when the researcher was not present. The results showed significant levels of self-management and significant increases in functional measures with some evidence for transfer into tasks of daily living. The current feasibility study suggests that DMIs can provide a valuable tool for the self-management of rehabilitation, providing motivational aids for long-term exercise.

Salivary biomarkers as objective data in music psychology research: Comparing live and recorded music to story readings

R. Orlando[1], C. Speelman[1], A. Wilkinson[2] and V. Gupta[3]
[1]Arts & Humanities, [2] Nursing & Midwifery, and [3] Medical & Health Sciences, Edith Cowan University, Perth, WA, Australia

Saliva contains many of the chemical markers found in blood. It provides a non-invasive, readily accessed and repeatable individual psychobiological snapshot. This exploratory study examined the comparative effects of live and recorded music on pain, anxiety and immune function of 50 tertiary students, 15 palliative care and 9 surgical patients. The paper describes the interventions as well as the saliva collection procedures and processes.

Participants, allocated quasi-randomly to the live, audiovisual, or audio conditions (n = 10), listened to 25 min of solo classical string music (reduced to 17 min for the clinical phase), or heard either a live or audio recorded reading of stories. The story readings provided controls for the music conditions. Saliva samples were taken before, immediately following, and 25 min after the interventions, to measure levels of cortisol, α-amylase, immunoglobulin-A, interleukin-1β, and salivary pH as indicators of pain, anxiety and immune function. Visual analogue scales for pain and anxiety and brief questionnaires supplemented the objective data.

It was anticipated that the live interventions would generate a greater reduction in measures of pain and anxiety, and greater improvement of immune function, than the audiovisual, which in turn would have a greater effect than the audio-only conditions. It was also predicted that music overall would have a larger effect on the outcome measures than stories. Results partly supported these hypotheses, in particular with the clinical population. Methodological issues and benefits of this psychoneuroendocrine approach to music psychology research will be discussed in light of the results.

Chanting Meditation Improves Mood and Social Cohesion

Gemma Perry,*[1] Vince Polito,#[2] William Forde Thompson*[3]

*Department of Psychology, Macquarie University, Australia

#Department of Cognitive Science, Macquarie University, Australia

[1]gemmaperry@outlook.com,[2]vince.polito@mq.edu.au,[3]bill.thompson@mq.edu.au

ABSTRACT

Chanting is a pervasive practice in almost every tradition all over the world. It has been found to improve attention and reduce depressive symptoms, stress and anxiety. The current study aimed to determine whether chanting "Om" for 10 minutes would improve attention, positive mood and increase feelings of social cohesion. The effects of vocal and silent chanting as a meditation practice were compared, as well as the effects of chanting for experienced and inexperienced chanters. It was hypothesized that vocal chanting would have a greater effect than silent chanting and experienced chanters would report stronger effects. Experienced and inexperienced chanters were randomly allocated to one of two conditions: vocal chanting or silent chanting. Prior to and following chanting, participants completed the *Digit-letter Substitution task,* the *Positive Affect Negative affect Schedule,* the *Multidimensional Measure of Empathy* and the *Adapted Self-Report Altruism Scale*. Following chanting participants also completed a *Social Connectedness Questionnaire* and a manipulation check. Results showed that positive affect and altruism increased more following vocal than silent chanting. Furthermore, whereas altruism increased following both vocal and silent chanting for experienced participants, it only increased following vocal chanting for inexperienced participants. No significant differences between vocal and silent conditions were observed for empathy, attention, or social connectedness. Overall, the results indicate that chanting has a positive effect on mood and social cognition. The findings are discussed in view of current understandings of the psychological and emotional effects of music and synchronization.

I. INTRODUCTION

Although chanting is a pervasive practice around the world, used in many traditions as a way of deepening spiritual awareness, there is very little understanding of the psychological, emotional and social implications of this widespread practice.

There are many different styles of chanting but all styles fall into two broad categories: vocalized and silent. Vocal chanting may be defined as the repetition of words or syllables that are either spoken or sung on the same note or a series of notes (Shearing, 2004). In contrast, "silent chanting" may be conceptualized as the repetition of imagined words or syllables in the absence of any vocalization.

As illustrated in Figure 1, vocal and silent chanting are forms of focused-attention (FA) meditation, a technique of concentration involving intense or prolonged focus on a single point (Bormann et al., 2006a). The most well known form of FA meditation is *mantra* meditation that involves focusing on the mental repetition of a specific sound or phrase, known as the mantra (Lutz et al., 2008; Bormann et al., 2014). Both vocal and silent chanting are forms of *mantra* meditation with the sound of phrase of concentration being the *mantra.*

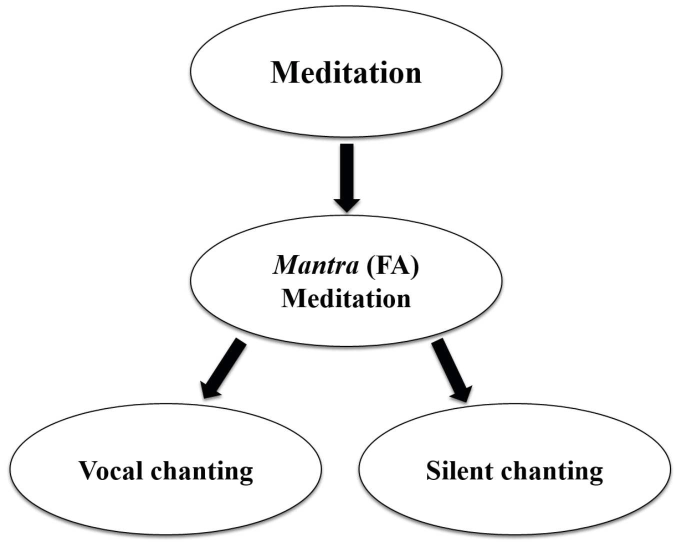


Figure 1. Chanting practices as a form of FA meditation.

Although a small body of research has identified some of the emotional and cognitive effects of chanting (Bernardi et al., 2001a; Kenny, Bernier & DeMartini, 2005; Pradhan & Derle, 2012; Wolf & Abell, 2003), no studies have compared the effects of silent and vocalized chanting, or the mediating effects of experience. The aim of the current study was to compare the effects of vocal and silent chanting on cognitive and affective states in both experienced and inexperienced chanters. Past research of explicit group synchronization has revealed that it leads to increased social cohesion (Valdesolo, Ouyang & DeSteno, 2010; Wiltermuth & Heath, 2009). Therefore, it is possible that explicit vocal chanting will enhance synchronization, such that the beneficial effects are stronger for this form of chanting than for silent chanting, which permits desychronisation given the absence of an explicit joint-action-task.

II. MATERIALS AND METHOD

A. Participants

The final analysis consisted of 45 inexperienced chanters (individuals who had chanted less than 5 times in total), consisting of 37 females and 8 males with ages ranging from 18 to 68 years (M = 25.11, SD = 13.07) and 27 experienced chanters (individuals engaging in chanting at least once a month for over 12 months), consisting of 13 females and 14 males with ages ranging from 18 to 68 years (M = 38.22, SD = 14.31). The inexperienced participants were recruited through the Macquarie University online participant pool or through social media. The experienced participants were recruited through social media and through a yoga studio. Participants recruited through Macquarie University were offered course credit for their participation. Members of the general public went into a draw to win 1 of 3 $50 vouchers for a yoga studio.

B. Materials

Participants completed the *Digit-letter Substitution task* (DLST; Natu & Agarwal, 1995), the *Positive Affect Negative affect Schedule* (PANAS), the *Multidimensional Measure of Empathy* (MME; Davis, 1980), and the *Adapted Self-Report Altruism Scale* (SRA; Rushton, Chrisjohn & Fekken, 1981) prior to the experimental phase of the research. Following the experimental phase, participants completed the above measures again and also completed the *Social Connectedness Questionnaire* (SCQ) and the manipulation check. The experimental phase of the research involved listening to a recording of chanting, which was either chanted along with in the vocal condition, or listened to in the silent condition. This recording can be found at:

https://www.youtube.com/watch?v=yoYrLM5rGX8&list=RDyoYrLM5rGX8#t=22.

C. Procedure

Experienced and inexperienced participants were randomly assigned to one of two experimental conditions: vocal chanting or silent chanting. Participants were invited to sit on a chair or cross-legged on a cushion on the floor. They first completed a consent form after which they completed the DLST, PANAS, MME and SRA. They were then instructed to maintain a straight spine whether on a chair or on the floor, close their eyes and chant the sound "Om" for 10 minutes either vocally or silently, timed to coincide with a recording. The duration of each repetition was 10 seconds, as used in previous studies (Bernardi et al., 2001b). After 10 minutes of chanting, all participants completed the DLST, PANAS, MME, SRA, SCQ, and the manipulation check.

III. RESULTS

A. Assumptions of Normality

Variables were examined for their conformity to the assumptions of a two-way between subjects analysis of variance (ANOVA). Assumptions of normality were met for all variables with the exception of negative affect, which had a positive skew and was slightly leptokurtic. It was deemed unnecessary to remove outliers as the dependent variables were normally distributed. Furthermore, the two-way ANOVA is robust to small violations of assumptions of normality (Kenny & Judd, 1986). Levene's tests were insignificant for all variables examined, indicating that homogeneity of variance was met. An alpha level of .05 was used for all significance tests.

B. Descriptive Statistics

The SCQ was completed post-chanting, and revealed that social connectedness was somewhat higher for experienced chanters in the silent condition, $M = 3.15$, $SD = 0.69$, than it was for experienced chanters in the vocal chanting condition, $M = 2.71$, $SD = 1.27$, inexperienced chanters in the silent condition, $M = 2.83$, $SD = 0.937$, or inexperienced chanters in the vocal condition, $M = 2.64$, SD = 0.953. For other measures, the mean and standard deviation of pre-chanting and post-chanting scores for each of the four conditions are displayed in Tables 1-4.

Table 1. Experienced participants' scores before and after silent chanting (N = 13).

Measure	Before *M* (*SD*)	After *M* (*SD*)	*t*	*p*
Attention**	60.92 (11.93)	65.23 (12.54)	-4.856	.000
Positive Affect	32.23 (6.59)	33.15 (9.13)	-0.746	.470
Negative Affect**	18.54 (2.11)	11.08 (1.55)	11.910	.000
Altruism**	50.92 (5.84)	53.08 (4.48)	-3.742	.003
Empathy	81.38 (7.43)	82.77 (6.44)	-1.133	.279

Note. Mean (*M*) and standard deviations (*SD*) (listed in parentheses) before and after chanting for measures of attention, positive affect, negative affect, altruism, and empathy.* ***p < .05, ** p < .01.***

Table 2. Experienced participants' scores before and after vocal chanting (N = 14).

Measure	Before *M* (*SD*)	After *M* (*SD*)	*t*	*p*
Attention	51.86 (14.72)	54.86 (12.25)	-1.580	.138
Positive Affect	30.57 (5.88)	29.93 (5.81)	0.374	.714
Negative Affect**	20.00 (4.26)	11.43 (2.98)	7.569	.000
Altruism**	50.07 (4.12)	55.07 (5.08)	-4.746	.000
Empathy	82.07 (6.73)	82.57 (7.31)	-0.559	.586

Note. Mean (*M*) and standard deviations (*SD*) (listed in parentheses) before and after chanting for measures of attention, positive affect, negative affect, altruism, and empathy.* ***p < .05, ** p < .01.***

Table 3. Inexperienced participants' scores before and after silent chanting (N = 23).

Measure	Before *M* (*SD*)	After *M* (*SD*)	*t*	*p*
Attention**	53.65 (12.83)	58.52 (12.85)	-2.963	.007
Positive Affect*	26.35 (7.74)	23.43 (8.75)	2.347	.028
Negative Affect**	18.87 (4.98)	11.26 (2.05)	8.020	.000
Altruism	49.52 (9.53)	50.13 (9.02)	-0.984	.336
Empathy	78.74 (7.00)	78.43 (8.30)	0.370	.715

Note. Mean (*M*) and standard deviations (*SD*) (listed in parentheses) before and after chanting for measures of attention, positive affect, negative affect, altruism, and empathy.* ***p < .05, ** p < .01.***

Table 4. Inexperienced participants' scores before and after vocal chanting (N = 22).

Measure	Before *M* (*SD*)	After *M* (*SD*)	*t*	*p*
Attention*	59.91 (9.66)	63.91 (12.31)	-2.202	.039
Positive Affect*	29.91 (6.24)	32.45 (8.77)	-2.096	.048
Negative Affect**	19.23 (2.98)	11.09 (1.74)	14.347	.000
Altruism**	47.55 (7.18)	50.91 (8.54)	-3.108	.005
Empathy*	79.27 (8.97)	81.45 (11.23)	-2.577	.018

Note. Mean (*M*) and standard deviations (*SD*) (listed in parentheses) before and after chanting for measures of attention, positive affect, negative affect, altruism, and empathy.* ***p < .05, ** p < .01.***

C. Analyses

Paired sample t-tests were used in order to examine the different effects of before and after chanting for each of the groups. Those measures that significantly increased or decreased following interventions are indicated in Table 1-4 by one ($p < 0.05$) or two ($p < 0.01$) asterisks. The change in score for attention, positive affect, negative affect and empathy were calculated by subtracting the 'pre score' from the 'post score', with the change in score used as the dependent variable. For example, the dependent variable for altruism was calculated by subtracting the pre-experiment altruism score from the post-experiment altruism score. As such, positive scores on the dependent variable represent a post-experiment increase in that variable, whereas a negative score represents a decrease in the variable. For each dependent variable a two-way ANOVA was carried out using the independent variables of engagement and experience.

D. Positive Affect

A two-way ANOVA using the mean difference of scores showed a significant interaction between type of chanting and amount of experience on positive affect, $F(3, 68) = 6.320$, $p = .014$, $\eta_p^2 = .085$. Positive affect increased more in the vocal chanting condition ($M = 1.00$, $SD = 1.00$) compared with the silent chanting condition ($M = 0.95$, $SD = 0.98$) and inexperienced chanters in the vocal condition ($M = 2.55$, $SD = 1.22$) showed a greater increase in positive affect than experienced chanters in the vocal condition ($M = -0.65$, $SD = 1.53$).

Post Hoc comparisons were conducted using Bonferroni adjusted alpha level of 0.025. These comparisons revealed that inexperienced chanters showed a significant overall increase in positive affect ($p = 0.003$) whereas experienced chanters showed no significant increase. Furthermore inexperienced chanters only showed a significant increase in positive affect for the vocal chanting condition ($M = 2.55$, $SD = 5.70$) and not the silent chanting condition ($M = -2.91$, $SD = 5.95$).

E. Altruism

A two-way ANOVA, using the mean difference scores showed a significant overall effect of type of chanting on altruism, $F(3,68) = 9.097$, $p = .004$, $\eta_p^2 = .118$. Altruism increased to a greater extent following vocal chanting ($M = 4.18$, $SD = 0.65$) than silent chanting ($M = 1.38$, $SD = 0.66$). Chanting experience did not significantly affect altruism, $F(3,68) = 2.935$, $p = .091$, $\eta_p^2 = .041$ and there was no significant interaction between level of experience and type of chanting on altruism $F(3,68) = 0.002$, $p = .961$, $\eta_p^2 = .000$.

F. Manipulation Check

To determine whether participants were chanting in both conditions, a manipulation check was conducted by measuring how much participants felt they *engaged* in the practice. The results of a two-way ANOVA revealed significantly higher scores in the vocal chanting compared with the silent chanting $F(3,68) = 31.451$, $p = .007$, $\eta_p^2 = .101$. Also, engagement scores were significantly higher for experienced chanters than for inexperienced chanters $F(3, 68) = 7.634$, $p = < .001$, $\eta_p^2 = .316$.

IV. DISCUSSION

Consistent with previous research, chanting increased positive mood, decreased negative mood and improved attention. Furthermore, altruism increased more following vocal chanting than silent chanting, suggesting that an explicit joint-action activity is more effective at creating feelings of social connectedness than a silent group activity. These findings indicate that chanting not only has a positive effect on mood and cognition; it also has the ability to improve feelings of social connectedness. Table 5 provides a synopsis of the significant increases and decreases of each variable following vocal and silent chanting. Significant benefits were observed in all four participant groups, but it is especially striking to note that *inexperienced* chanters exhibited significant benefits following vocal chanting in all five measures of mood and social cohesion. Although the source of these effects has yet to be determined, Figure 2 illustrates some hypothetical mechanisms underlying these benefits.

Table 5. Significant increases and decreases of measures for experienced and inexperienced participants following silent and vocal chanting conditions.

	Attention	Positive Affect	Negative Affect	Altruism	Empathy
Experienced Silent	↑	-	↓	↑	-
Experienced Vocal	-	-	↓	↑	-
Inexperienced Silent	↑	↓	↓	-	-
Inexperienced Vocal	↑	↑	↓	↑	↑

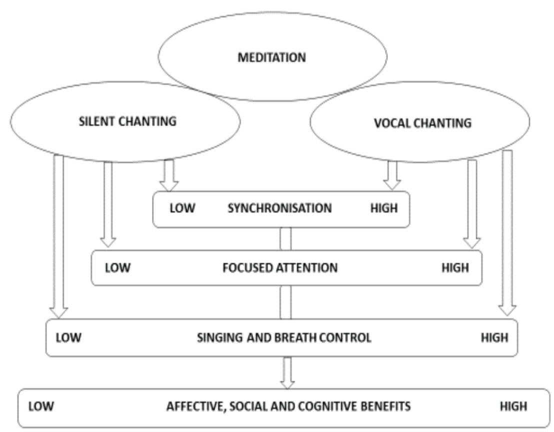


Figure 2. Factors associated with the benefits of chanting.

V. CONCLUSION

Chanting is a pervasive practice that has been part of human behaviour for thousands of years, yet the emotional and cognitive effects of chanting are not well understood. The results of this investigation demonstrate that the benefits of vocal chanting may be mediated by three factors: group synchronization, which increases feelings of social cohesion; physiological changes (breath control and singing), which may contribute to increased positive mood; and focused attention, which may inhibit ruminative thinking and lead to increased positive mood.

As one of the first studies to systematically investigate the benefits of chanting, the current study provides a basis for future research on the emotional, social, and cognitive benefits of this pervasive human activity. As revealed by this investigation, chanting has emotional and social benefits for both experienced and inexperienced individuals, and these benefits can be observed following either vocal or silent chanting. The extent of these benefits was dependent on the type of chanting meditation and the experience of participants, with inexperienced participants who engaged in vocal chanting reaping the greatest number of significant benefits. Such results are encouraging, and suggest that chanting may be an effective tool for enhancing mood and creating a sense of social cohesion. These benefits, in turn, may be associated with increased health and wellbeing.

ACKNOWLEDGMENTS

This research was supported by an Australian Research Council Discovery Grant DP130101084 awarded to WFT and by a fellowship awarded to VP by the Australian Research Council Centre of Excellence in Cognition and its Disorders.

REFERENCES

Bernardi, L., Gabutti, A., Porta, C., & Spicuzza, L. (2001a). Slow breathing reduces chemoreflex response to hypoxia and hypercapnia, and increases baroreflex sensitivity. *Journal of Hypertension*, 19(12), 2221-2229.

Bernardi, L., Sleight, P., Bandinelli, G., Cencetti, S., Fattorini, L., Wdowczyc-Szulc, J., & Lagi, A. (2001b). Effect of rosary prayer and yoga mantras on autonomic cardiovascular rhythms: Comparative study. *British Medical Journal*, 323(7327), 1446-1449.

Bormann, J. E., Becker, S., Gershwin, M., Kelly, A., Pada, L., Smith, T. L., & Gifford, A. L. (2006a). Relationship of frequent mantra repetition to emotional and spiritual wellbeing in healthcare workers. *Journal of Continuing Education in Nursing*, 37(5), 218224.

Bormann, J. E., Oman, D., Walter, K. H., & Johnson, B. D. (2014). Mindful attention increases and mediates psychological outcomes following mantra repetition practice in Veterans with posttraumatic stress disorder. *Medical Care*, 52, S13-S18.

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. Catalog of Selected Documents in Psychology. 10(85). Retrieved. from http://www.researchgate.net/publication/34891073_A_Multidimensional_Approach_to_Individual_Differences_in_Empathy

Kenny, M., Bernier, R., & DeMartini, C. (2005). Chant and be happy: The effects of chanting on respiratory function and general wellbeing in individuals diagnosed with depression. *International Journal of Yoga Therapy*, 15, 61-64.

Lutz, A., Slagter, H. A., Dunne, J. D., & Davidson, R. J. (2008). Attention regulation and monitoring in meditation. *Trends in Cognitive Sciences*, 12(4), 163-169.

Natu, M. V., & Agarwal, A. K. (1995). Digit letter substitution test (DLST) as an alternative to digit symbol substitution test (DSST). Human Psychopharmacology: Clinical and Experimental, 10, 339-343. doi: 10.1002/hup.470100414

Pradhan, B., & Derle, S. G. (2012). Comparison of effect of Gayatri Mantra and Poem Chanting on Digit Letter Substitution Task. *Ancient Science of Life*, 32, 89–92.

Rushton, J. P., Chrisjohn, R. D., & Fekken, G. C. (1981). The altruistic personality and the self-report altruism scale. Personality and Individual Differences, 2(4), 293-302. Retrieved from http://www.subjectpool.com/ed_teach/y3project/Rushton1981THE_ALTRUISTIC_PERSONALITY.pdf

Shearing, C. R. (2004). *Gregorian chants: An illustrated history of religious chanting*. London: Mercury Books.

Wiltermuth, S. S., & Heath, C. (2009). Synchrony and cooperation. Psychological Science, 20(1), 1-5. doi: 10.1111/j.1467-9280.2008.02253.x

Wolf, D. B., & Abell, N. (2003). Examining the effects of meditation techniques on psychosocial functioning. *Research on Social Work Practice*, 13, 27-42.

Valdesolo, P., Ouyang, J., & DeSteno, D. (2010). The rhythm of joint action: Synchrony promotes cooperative ability. Journal of Experimental Social Psychology, 46(4), 693695. doi: 10.1016/j.jesp.2010.03.004

Advancing Interdisciplinary Research in Singing: Group singing and wellbeing

Jacques Launay[1]
[1]Psychology Division, Brunel University London, London, UK

In recent years choirs have become increasingly popular, and this is likely due to the effects that they can have on people's wellbeing, and the capacity they have for rebuilding a sense of community. This symposium brings together researchers who are investigating the positive role choirs and group singing can play in a variety of contexts. This includes the role of choirs on intercultural bonding on college campuses, in people with Parkinson's Disease, in building communities in London, and in older adults taking group singing lessons. This is relevant to anyone with an interest in the ways that community and health can be influenced by singing in groups, and in particular those who are interested in the ways in which singing can produce these strongly positive effects. Presenters will discuss the impact that their choral groups have had on the health, social and intellectual lives of singers, as well as the mechanisms that might underpin these changes. Challenges will also be discussed. Under the auspices of the AIRS (Advancing Interdisciplinary Research in Singing) project, presentations will primarily address ongoing work, but will also outline prospective projects that aim to expand beyond the populations currently studied. Daniel Weinstein and his colleagues (UK) explore the effect of choral singing on the fostering of social closeness. In particular they considered whether increasing size of the group is a limiting factor and found that this was not the case. Frank Russo (Canada) describes the formation of a choral group for persons with Parkinson's disease. Jennifer Bugos (US) reports on a study in which older adults engaged in group singing lessons for a period of six weeks. Annabel Cohen and colleagues (Canada/Malta) describe the challenges of creating a multicultural choir on a college campus. Jacques Launay (UK) will discuss the papers from the point of view of social bonding and health benefits but also from the perspective of connecting research to practice, helping to answer the question of whether our research knowledge can benefit the public engaged in singing activities and how group singing can inform understanding of the power of musical shared engagement.

Changes in social connectedness as a result of community singing in small and large choirs

Daniel Weinstein[1], Jacques Launay[2], Eiluned Pearce[3], Robin I.M. Dunbar[3,5], Lauren Stewart[1,4]
[1]Goldsmiths, University of London, Lewisham Way, New Cross, London, SE14 6NW, UK
[2]Brunel University of London, Kingston Ln, Uxbridge, Middlesex UB8 3PH, UK
[3]Department of Experimental Psychology, University of Oxford, Tinbergen Building, Oxford, OX1 3UD, UK
[4]Center for Music in the Brain, Department of Clinical Medicine, Aarhus University & The Royal Academy of Music Aarhus/Aalborg, Denmark
[5]Department of Computer Sciences, Aalto University, Espoo, Finland

Prior research has shown that singing has many positive physical health benefits as well as can foster prosocial behavior. We investigated group singing as a means of social bonding in both small and large choir contexts. While previous research has shown that music making in small group contexts is an effective means of maintaining social cohesion, the question of whether this effect could occur in larger groups remains unanswered. We set out to compare the social bonding effects of singing in small, familiar contexts versus a large, unfamiliar context. The current study recruited individuals from a community choir, Popchoir, that meets in small choirs weekly, and comes together annually to form a large choir combining individuals from the smaller subchoirs. Participants gave self-report measures of affect and social bonding and had pain threshold measurements taken (as a proxy for endorphin release) before and after 90 minutes of singing. Positive affect, social bonding measures, and pain thresholds all increased across singing rehearsals. We found no significant difference for pain thresholds between the small versus large group context. Additionally, levels of social bonding were found to be greater at pre- and post-levels for the small (familiar) choir context. However, the large choir context showed a greater change in social bonding survey measures as compared to the small context. We find that group singing is an effective means to foster social closeness, even in large group contexts where many individuals are unfamiliar with one another.

Effects of choral singing on facial and vocal expressiveness in people living with Parkinson's Disease

Frank A. Russo[1], Esztella Vezer[1], and Steven R. Livingstone[1]
[1]Department of Psychology, Ryerson University, Toronto, Ontario, Canada

People with Parkinson's disease (PD) experience difficulty producing expressive aspects of vocal and facial communication, resulting in a permanently flat expression that has been referred to as "masked face syndrome". It has also been found that PD patients have decreased ability to correctly identify emotion in the facial expressions of others (e.g., Dujardin et al., 2004; Buxton, McDonald & Tippett, 2013). Identification deficits appear to be due in part to deficits in automatic facial mimicry (Livingstone, Vezer, McGarry, Lang & Russo, under review). Because expressive aspects of vocal and facial communication are an important part of social interaction, impairments can lead to an array of emotional health issues such as anxiety, depression and reduced quality of life (Cummings, 1992; McDonald, Richard & DeLong, 2003). Despite its importance for social and emotional wellbeing, few rehabilitative programs have targeted expressive aspects of vocal and facial communication in PD patients. Based on recent research that our lab has conducted demonstrating a positive effect of singing on expressive aspects of communication in PD, the current study reports on a 13-week choir for people with PD that emphasizes vocal and facial expressiveness. We hypothesize that the training program will improve participants' expressiveness as well as their emotional mimicry and their ability to accurately perceive emotion in others, ultimately enhancing interpersonal communication and quality of life.

Vocal Improvisation and Verbal Fluency in Older Adults Following Group vocal lessons

Jennifer A. Bugos[1,2], Chloe Kantoris[1]
[1]School of Music, University of South Florida, Tampa, Florida, USA
[2]Center for Music Education Research, University of South Florida, Tampa, FL, USA

Music training can contribute to melodic expectations. Research suggests that vocal improvisations by young adult musicians end on the tonic, and those by young adult non-musicians are less likely to do so (Pan & Cohen, 2015). Little is known regarding the effects of training on vocal improvisations in novice older adults. Vocal improvisation requires the ability to generate musical phrases, and similarly, verbal fluency is the ability to generate words. Deficits in verbal fluency are one of the first symptoms of cognitive impairment in aging adults (Laisney et al, 2009). Thus, we examined these areas in conjunction with processing speed, an area known to affect verbal fluency. The purpose of this study was to evaluate vocal improvisations and cognitive performance in novice older adult vocal students following group vocal lessons. Thirteen healthy, older adults were recruited from independent living facilities, and screened for cognitive impairment. Participants completed measures of verbal fluency, simple processing speed, music aptitude, and vocal achievement pre- and post-training. Participants attended eight weeks of group vocal instruction. Each two-hour class included exercises in vocal technique, music reading, and vocal independence. Vocal improvisation exercises were performed with better intonation post-training; however, most participants repeated the same material multiple times prior to ending the phrase. Consistent melodic contour for improvised endings was found pre- and post-training. In addition, familiar melodies were represented in repeated material. Results of a pair-samples *t*-test showed significant increases in letter fluency, post-training. Group vocal training can instill confidence in novice vocalists allowing them to improvise more freely. Improvisation exercises within the context of a vocal training course may contribute to enhanced verbal fluency, an area that can impact the quality of life in older adults.

Benefits of Forming a Multicultural Choir and Song Circle on a University of Campus

Annabel J. Cohen[1], Godfrey Baldacchino[1,2], Karen Ludke[3], Liliya Nafikova[1], Bing-Yi Pan[1]
[1]University of Prince Edward Island, Prince Edward Island, Canada
[2]University of Malta, Malta
[3]Edge Hill University, UK

Establishing social bonds is one of the first challenges students face on arriving at a university for the first time, and this challenge is greater for those who "come from away". Joining a choir might provide one means to meet this challenge, however, many campus choirs are exclusive audition choirs, and their aim is often to perfect a performance program. In contrast, a song circle approach might have wider appeal and offer common ground for friendship and appreciation of one's own and others' cultures. Over a five-year period a protocol has been in development for a campus "multicultural choir and song circle" in which participants are encouraged to share in the responsibility for the repertoire while they are provided the scaffolding to lead with the help of a somewhat "behind-the-scenes" team consisting of the nominal choir director, collaborative pianist, and research director and associates. A questionnaire before and after a 10-week session revealed that the goals of the 19 participants in joining were met and in some cases were exceeded. While sharing the over-riding goal to have fun, diverse interests of members such as singing to perform vs singing to share in joint activity can sometimes conflict. Pressure to meet standards for external performance requests can upset the group cohesion. A solution entailing two groups – one that performs and one that does not—is not likely practical, at least on a small campus, and ways of achieving compromise will be discussed. The concept of the multicultural choir and song circle initiated in the small Canadian University has been duplicated at a larger European University for two consecutive years suggesting the transferability of the concept. The research suggests that the university multicultural choir and song circle adds social capital to the parent university as well as to the broader community.

Music Training and Neuroplasticity: Cognitive Rehabilitation of Traumatic Brain Injury Patients

Berit Marie Vik[1*], Geir Olve Skeie[2,3], Karsten Specht[1,4]

[1]*Department of Biological and Medical Psychology, University of Bergen, Norway*
[2]*Department of Neurology, Haukeland University Hospital Bergen, Bergen, Norway*
[3]*Grieg Academy Research Centre of Music Therapy (GAMUT), University of Bergen, Bergen, Norway*
[4]*Department of Medical Engineering, Haukeland University Hospital Bergen, Bergen, Norway*
[*]vikmusic@online.no

ABSTRACT

Musical training has become a useful framework to study brain plasticity and to be a possible tool in neurologic rehabilitation. In this study we explored the effects of playing the piano on traumatic brain injury (TBI) patients with cognitive impairment. We addressed the question if this approach would stimulate neural networks in re-routing neural connections and link up cortical circuits that had been functional inhibited due to diffuse axonal injury (DAI) and thus restore cognitive function. Here, for the first time we demonstrated structural and functional brain changes after 8 weeks of musical training on TBI patients having attention, memory and social interaction problems. 3 groups participated in the study, one TBI group (*n*=7), two groups of healthy participants (*n*=11), one with music training and one baseline group (*n*=11) without music. Functional neuroimaging scans (fMRI) and neuropsychological tests were performed pre-post intervention. Analyzes revealed training-related plasticity in the two groups with music, significant change in Orbitofrontal cortex (OC) controlling learning, attention, memorization and social interaction. fMRI results fits well with outcome from neuropsychological tests, significant enhancement of cognitive performance in both music groups. Interviews pre-post in the TBI group documented that six out of seven participants were able to return to their jobs after the music intervention period. The key findings of this study are the clear evidence for a causal relationship between musical training and reorganization of neural networks promoting enhanced cognitive performance. These results adds a novel music-supported intervention within rehabilitation of TBI patients with cognitive deficits.

Keywords: Music perception and cognition, neuroplasticity,TBI,rehabilitation, fMRI

I. INTRODUCTION

Musical training is a highly multimodal complex stimuli for the brain. Reading and playing an instrument simultaneously receives and transmits visual (music literacy), auditory (listening), and kineasthetic (motor) information to a specialized brain network (Thaut 2010). Musical practice over time has shown to effectively change the structure and enhance the function of many brain areas (Herholz and Zatorre 2012). Music making activities are powerful tools to engage multisensory and motor network, induce changes within these networks, and foster links between distant,but functionally related brain regions with musical practice. These multimodal effects of music together with music's ability to engage emotion and reward system in the brain can facilitate and enhance therapeutic approaches aimed towards rehabilitation from neurological and psychiatric disorder (Stewart et al.2003;Schlaug et al.2009;Särkamö et al.2013;Hegde 2014;Altenmüller and Schlaug 2015;Herholz et al.2015).The intervention method presented is categorized within Neurologic Music Therapy (NMT) defined as «the therapeutic application of music to cognitive, sensory, and motor dysfunction due to neurologic disease of the human nervous system (Thaut 1999:2). The NMT method is based on neuroscience models of music perception and the influence of music on changes in non-music related brain areas.

Neural cross-activity in the neural networks during music performance activates almost all regions of the brain (Tramo 2001;Koelsch et.al. 2005; Zatorre 2005; Parsons et al.2005; Zatorre et al.2007). Extensive research within music perception and cognition reveals shared networks language/music (Patel et al.1998, Parsons et al.2005;Brown et al.2006; Koelsch 2012). Together with the effecs of association/priming/mnemonic neural activity stimulating episodic and semantic memory networks during musical practice, we suggest this neural cross-activity may generate new neural connections. The main objective of the designed music-supported intervention in the present study is to restore cognitive performance in TBI patients.

There is little knowledge about the potential neuroplastic changes induced by musical training in cognitive rehabilitation of TBI patients (Hegde 2014). Traumatic brain injury is defined as brain injury caused by external trauma. TBI has been recognized as a major public health problem worldwide ans is one of the common causes of disability. The economic impact of mild TBI injuries account for 44% of the annual cost of TBI in the United States (Balanger et al.2007).

The present research field comprises domains of music psychology, cognitive neuroscience of music, neuroscience, musicology emphasizing training related brain plasticity and the clinical application of music.

II. METHODS AND PROCEDURES

Participants

Three groups were recruited for the study. One group of 7 mild TBI out-patients two years post injury(to control for the variable of self-recovering within the two first years) with music intervention, one control group of 11 healthy participants with music intervention and one baseline control group of 11 healthy participants without music.

Study design

We applied a between-group design as well as a longitudinal within-subject design. All participants from the TBI and the first control group were examined pre- and post intervention. The second control group was also examined twice, but without any intervention in between. During the pre- and post intervention assessment, all participants were examined with a neurospychological test-battery, functional and structural MRI brain scanning with two experimental fMRI paradigms, resting state fMRI, and diffusion tensor imaging (DTI). Further, semi-structured interviews pre-post were conducted in the TBI group.

Piano Training Protocol

8 weeks piano-tuition intervention program was developed. All participants were non-musicians. The participants received two lessons a week in the lab with instructor, 30 minutes each lesson with instructions to practice a minimum of 15 minutes each day. Mean training time per week was 3 hours. 20 short pieces for beginners was curriculum, the last pieces for two hands, melody in right hand and chords in left.

Neuropsychological tests

A test battery of three different neuropsychological tests was applied during pre-and post examinations of the participants. The test battery contained : Mini Mental Status Test (MMS), California Verbal Learning Test (CVLT) assessing learning strategies, memorization and retrieval of information, and Stroop Word/Color Test for assessing reading speed and attention and the ability to switch attention.

fMRI Scanning Session

Functional and structural fMRI scans were collected pre-post intervention for all three groups. Besides a high-resolution T1-weighted scan, two task related fMRI sessions were included in the scanning protocol, where one task was a pitch-discrimination task and the other one was Tonika-Dominant-Tonika chords, from extracts of Western music cadence pattern.

Results

Neuropsychological tests revealed highly significant effect of musical training on executive funcions related to attention, learning strategies and retrieval of memories, fig.2. MMS test has been excluded from the study because of a ceiling effect. Stroop word/color test, fig.1, is a measure of reading speed and attention, but has not the factor of learning strategies and memory retrieval as CVLT. There were no significant main effects of interactions. There was a trend of a repetition effect in the two control groups.

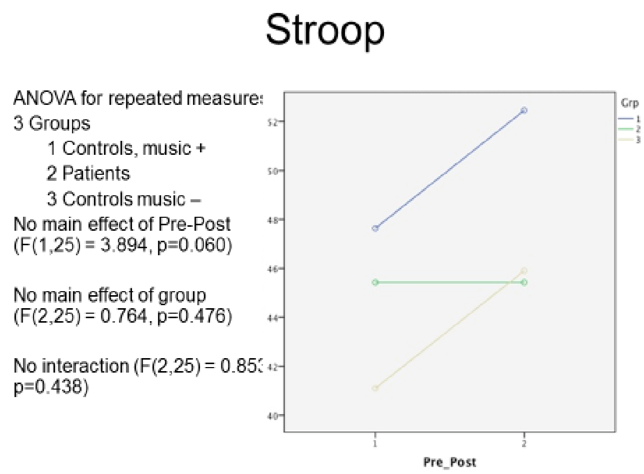


Fig.1. ANOVA for repeated measures shows no significant interaction of the groups. Effect only in the two control groups.No effect in the patient group.

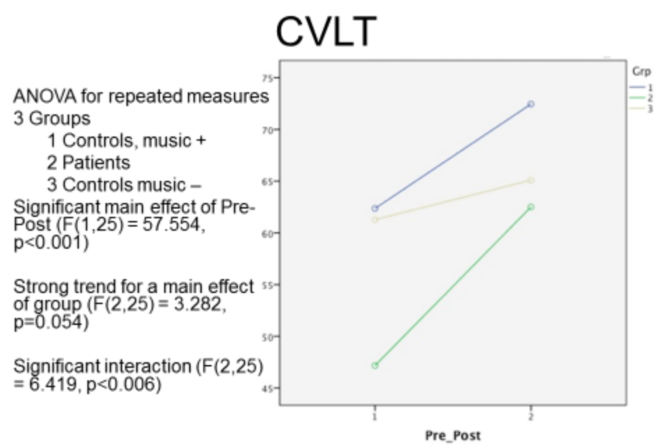


Fig.2. Significant main effect of Pre-Post (p<0.001). Significant interaction between groups (p < 0.006). Patient group increases cognitive performance up to the level of healthy controls pre-intervention.

fMRI

Tonika-Dominant-Tonika task

We found significant functional changes in orbitofrontal cortex in the patient group, no significant effect in control-group with music. Fig.3.

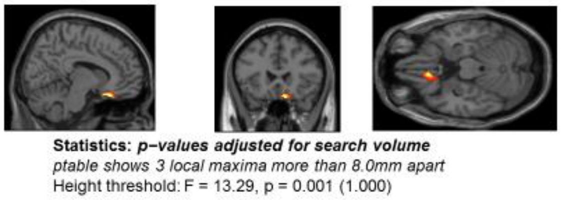


Fig 3. Training effects of eight-week piano training. Functional activation changes in Orbitofrontal cortex patient group.

III. DISCUSSION

In analysing fMRI data we found significant changes in orbitofrontal cortex. This networks regulates higher order cognitive processing such as executive functions including attention, concentration, decision-making, impulse control and, social interaction. The OC has a widely distributed interconnected neural networks to almost all areas of the brain. Auditory processing and motor areas are densely interconnected with other prefrontal cortical regions, reflecting integration for executive motor control which is evident in playing the piano. OC's connectivity to association corties are of special interest as this is one of the most important connectivities in developing new pathways in reference to the NMT intervention applied (Schlaug 2009:197-2008). This neural process may link up new pathways and explain the functional and structural changes in the neural networks as reported in this study.Interestingly, there were significant differences between the two music groups in structural changes post intervention. In the control group with music there was a significantly increased synaptic density in Basal Ganglia (BG), the motor area. There were no BG changes in the TBI patient group. However, this group revealed significant changes in Lingual Gyrus, Occipital cortex, the neural networks for sight-reading musical notation. There may be several reasons for this difference. Firstly, the control group may employ less effort in reading the notation, while the patient group concentrated more in learning and reading notes. This area is active in analysing logical conditions and recognition of words, which some TBI patients reported as their spesific deficit. Secondly, the control group had a larger repertoire, and therefore employed more training/time in stimulating Basal Ganglia. The training time may be responsible for the differences in structural changes in BG and explain why the patient group did not have any significant changes in the motor area.

Another possible theory of the structural changes in the TBI group's Lingual Gyrys (LG) and changes in Orbitofrontal cortex, may be the connectivity between the Temporal Pole (TP) and Ventral Visual Stream during social interaction. A recent study demonstrated that the TP integrates information from different modalities (in this study music) and top-down modulates lower-level perceptual areas in the ventral visual stream as a function of integration demands (Pehrs et al.2015). TP is part of the association cortex and involved in multimodal sensory integration (Olson et al.2007;Skipper et al.2011). The results indicates interactive neural activity within the stream consistent with finding from object recognition. The authors suggests there may be a TP bottom/up perceptual information to higher/order cognitive areas. For example by functional connectivity to the orbitofrontal cortex (Kahnt et al.2012). We may suggest a possible explanation of the TBI groups structural change in Lingual Gyrus and Orbitofrontal cortex correlating with the above theory, namely a bottom-up connectivity from TP to OC responsible for functional and structural changes in the neural networks.

Limitations to the study is the spare number of participants in the TBI patient group. Future studies should also include a second TBI patient group without music intervention as control to the TBI group with music. However, as conventional therapies already had been applied within the participants in the TBI group, we found the present methodology to be sufficient for the study.

IV CONCLUSION

The novel key finding of this study is that regular piano playing during 8 weeks can lead to structural and functional reorganization of the neural networks in TBI patients with cognitive impairment. The results demonstrated significant training-related neuroplasticity with enhanced cognitive performance in both music groups, supporting the aim of the intervention, to restore cognitive function in TBI patients. Additional research may consider the variable of intervention length and practising time, together with increased number of participants,to establish an intervention adapted for future cognitive music supported rehabilitation programs for TBI patients.

REFERENCES

Altenmüller, E. & Schlaug, G. (2015). Apollo's gift: new aspects of neurologic music therapy. *Progress in Brain Research.* Volum 217, ISSN 0079-6123.

Balanger, H.G., Vanderploeg, R.D., Curtiss, G. & Warden, D.L. (2007). Recent neuroimaging techniques in Mild Traumatic Brain Injury. Journal of Neuropsychiatry and Clinical Neurosciences 19,5-20. In Sigurdadottir, S. (2010). A prospective study of Traumatic Brain Injury. *Dissertation PhD.* Department of Psychology. Faculty of Social Sciences. University of Oslo, Oslo, Norway.

Brown, S., Martinez, M.J. & Parsons, L.M. (2006). Music and Language side by side in the brain.: a PET study of the generation of melodies and sentences. *European Journal of Neuroscience,* Vol.23, pp.2791-2803.

Hegde, S. (2014). Music-based cognitive remediation therapy for patients with traumatic brain injury. *http://journal Frontiersin.org/article/*10.3389/fneur.2014.00034/full.

Herholz, S.C., Zatorre, R.J. (2012). Musical training as a framework for brain plasticity: behaviour, function, and structure. *Neuron* 76:486-502. Doi: 10.1016/j.neuron.2012.10.011.

Herholz, S.C., Coffey, E.B.J., Pantev, C. & Zatorre, R.J. (2015). Dissociation of Neural Networks for Predisposition and for Training-Related Plasticity in Auditory-Motor Learning. *Cerebral Cortex,* 2015, 1-10.

Kahnt, T., Chang, L.J. Park, S.Q., Heinzle, J. & Haynes, J.D. (2012). Connectivity-based parcellation of the human orbitofrontal cortex. *Journal of Neuroscience,* 32:6240-6250.

Koelsch, S., Fritz, T., Schulze, K., Alsop, D. & Schlaug, G. (2005). Adults and children processing music. An fMRI study. *Neuroimage.*25 (2005:1068-1076).

Koelsch, S. (2012). Response to target article "Language, music, and the brain: a resource-sharing framework". In Koelsch, S. *Language and Music as Cognitive Systems.*Oxford: Oxford University Press.

Olson, I.R., Plotzker, A., Ezzyat, Y. (2007). The enigmatic temporal pole: a review of findings on social and emotional processing. *Brain.* 130:1718-1731.

Parsons, L. Sergent, J., Hodges, D.A., & Fox, P.T. (2005). The brain basis of piano performance. *Neuro-Psychologia.* 43 (2005) 199:215.

Patel, A.D., Peretz, I., Tramo, M., & Labreque, R. (1998). Processing prosodic and musical patterns: a neuropsychological investigation. Brain Lang., 61, 123-144. In *Music and language side by side in the brain: A PET study of the generation of melodies and sentences.European Journal of Neuroscience.* Vol. 23, pp.2791-2803.

Schlaug, G. (2009). Music, musicians, and brain plasticity. In *The Handbook of Music Psychology.* Oxford: Oxford University Press.

Schlaug, G. (2009). Listening to and Making Music Facilitates Bain Recovery Processes. In *The Neuroscience and Music III. Disorders and Plasticity.* Boston: Blackwell Publishing on behalf of the New York Academy of Sciences.

Skipper, L.M., Ross, L.A., & Olson, I.R. (2011). Sensory and category subdivision within the anterior temporal lobes. *Neuropsychologia.* 49:3419-3429.

Stewart, L., Henson, R., Kampe, K., Turner, R., and Frith, U. (2003). Brain changes after learning to read and play music. *Neuroimage,*20, pp. 71-83.

Särkamö, T., Tervaniemi, M. & Huotilainen, M. (2013). Music perception and cognition: development, neural basis, and rehabilitation use of music. *WIREs Cogn. Sci2013.* Doi: 10.1002/wcs.1237.

Thaut, M., (1999). Training Manual for Neurologic Music Therapy.

Thaut, M., (2010). Neurologic Music Therapy. Therapy in Cognitive Rehabilitation. *Music Perception.* Volume 27.

Tramo, M., (2001). Music of the Hemispheres. *Science Magazine.*5. January 2001 pp.54.56.

Zatorre, R., (2007). Neural Specialization for Tonal Processing. *The Cognitive Neuroscience of Music.* Oxford: Oxford University Press.

Help Musicians UK Hearing Survey: Musicians' hearing and hearing protection

Alinka Greasley[1], Robert Fulford[1], Nigel Hamilton[2] and Maddy Pickard[2]
[1]School of Music, University of Leeds, Leeds, UK
[2]Help Musicians UK, London, UK

The health of musicians' hearing has received attention due to increased legislation protecting employees from noise in working environments. Help Musicians UK (HMUK) is a charity for professional musicians of all genres throughout their careers. In 2014, HMUK conducted a national survey into musicians' health, and found that 47% of the sample (n=552) reported that they had experienced hearing loss. In response, HMUK conducted another survey investigating the prevalence and type of hearing problems, help-seeking behaviour, worry about noise at work, and use of hearing protection. A sample of 693 professional musicians took part, the majority being orchestral or instrumental musicians. Quantitative and qualitative data were analysed using SPSS and NVivo software. Results revealed contrasting patterns of associations between subjective perceptions of risk, experiencing symptoms, help seeking behaviour, hearing loss (HL) diagnosis, worry about noise at work, awareness of noise legislation and knowing colleagues with hearing loss. A large proportion of the sample had experienced HL or other hearing issues and many attributed hearing problems to their musical careers. Whilst two-thirds reported wearing hearing protection, frequently used, inexpensive forms of protection were given much lower effectiveness ratings than infrequently used expensive forms of protection. Orchestral and instrumental musicians were significantly more likely to use protection than singers and pianists. No association was found between having a test and the use of protection, suggesting that hearing tests should be accompanied by advice about the risks of NIHL to support improved protection behaviour and uptake. Results will be interpreted in relation to existing literature on musicians hearing and health promotion behaviour, and the data used to delineate the potential roles and responsibilities of employers, musicians' charities and musicians themselves in improving hearing protection uptake.

Predicting emotional well-being: The roles of emotional sensitivity to music and emotion regulation using music

Tan-Chyuan Chin[1] and Nikki Rickard[1,2]
[1]Melbourne Graduate School of Education, The University of Melbourne, Victoria, Australia
[2]School of Psychological Sciences, Monash University, Victoria, Australia

Music engagement has been associated with a wide range of health and well-being benefits. There are however differential outcomes depending on a variety of factors such as listener contexts and emotion regulation styles. This study explored the relationships between the various motivations for engaging with music and self-reported perceptions of musical capacity as captured by the MUSEBAQ, a comprehensive and modular instrument for assessing musical engagement, in a sample of 2964 adults (aged 18-87; 59% females; 55% without formal music training). This study further examined the mediating role of emotional sensitivity to music between emotion regulation using music and emotional well-being. Findings from mediation analyses suggest that the purposeful use of music to regulate emotions may not necessarily predict better well-being outcomes. Higher levels of emotional well-being were predicted only with a greater capacity to experience strong emotions through music. This study suggests that when evaluating benefits of music engagement on well-being outcomes, individual variations in music capacity and engagement need to be carefully considered and further researched.

Efficacy of a self-chosen music listening intervention in regulating induced negative affect: A randomised controlled trial

Jenny M. Groarke[1], Michael J. Hogan[1], Phoebe McKenna-Plumley[1]
[1]School of Psychology, National University of Ireland, Galway, Galway, Ireland

Background
Affect regulation is a common function of music listening. However, experimental findings regarding its efficacy are mixed. The majority of previous studies have employed university-aged samples, a silent control condition, and researcher-prescribed classical, or relaxing music.

Aims
The current study evaluates the efficacy of a self-chosen music listening intervention with older (OA) and younger adults (YA).

Method
Forty YA (18-30 years, $M = 19.75$, $SD = 2.57$, 14 males) and forty OA (60- 81 years, $M = 69.06$, $SD = 5.83$, 21 males) completed a measure of music listening functions and training, demographic information, and were asked to make a playlist with 15 minutes of music they would listen to in a stressful situation. Eligible participants visited the laboratory and were randomised to either the intervention (10 minutes of music listening) or the control condition (10 minutes of listening to a radio documentary). Negative affect (NA) was induced in all participants using the Trier Social Stress Test, followed by the intervention/control condition. Measures of self-reported affect (Visual Analogue Scales: *Excited-Depressed, Happy-Sad, Calm-Tense, Content-Upset, Relaxed-Nervous, Serene-Stressed, Alert-Bored, Active-Fatigued*) were taken at baseline, post-induction and post-intervention.

Results
A 3 (baseline, post-induction, post-intervention) x 2 (control, intervention) x 2 (younger, older adult) mixed-ANOVA found significant time*group interaction effects for *Happy-Sad, Relaxed-Nervous, Active-Fatigue,* and time*group, and time*age interactions for *Excited-Depressed* and *Alert-Bored*. A self-chosen music listening intervention was more effective than an active control at reducing induced NA, and was most effective for older adults - reducing NA below baseline levels.

Conclusions
OA have rarely been included in music studies. Findings here suggest music listening may support advanced emotion regulation capabilities in older adulthood, and will be discussed in relation to *socio-emotional selectivity theory*. Age differences in intervention effects may also result from differences in music chosen for the purposes of affect regulation. 66% of the OA sample chose classical music pieces, similar to those used in previous research, however only 3% of the YA sample included classical music, with more choosing upbeat music from rock, pop, and dance genres. Results from this self-chosen music listening intervention are in line with past studies using classical pieces or

music to induce particular moods. We found that self-chosen music for affect regulation significantly reduced negative affect, particularly in older adults for whom it decreased below baseline levels, in comparison to the control condition.

Music in clinical settings

Nora K. Schaal[1]
[1]Department of Experimental Psychology, Heinrich-Heine-University, Düsseldorf, Germany

Music is a complex auditory stimulus and a universal feature of human societies. The use and effects of music in clinical settings is of growing interest. Studies have shown that listening to music before or during a surgery can reduce patients' anxiety and pain levels and that music therapy can add beneficial value in health care and rehabilitation. Research in this area has recently become more robust by combining more diverse disciplines but there is a clear need of research looking at the complexity of music perception and the extent and limits of the treatment effects of music. There is a demand for research looking at the specificities of music and the mechanisms of how these influence wellbeing and mood. Additionally, small sample sizes and the lack of studies combining subjective and objective outcome measures evaluating anxiety and pain levels or benefits in health care restrict the significance of reported effects. Therefore, the present symposium combines fundamental research and applied science in order to highlight new findings on the perception of music, the biological mechanisms of the effects of music and demonstrates how music can be used in clinical settings. More precisely, the exploration of the neurochemistry and social flow of singing will be examined and music perception in individuals with cochlear implants will be discussed. Further, the symposium includes a paper presenting clinical data investigating the effects of listening to music during a caesarean section on patients' subjective and objective anxiety levels. The aim of the symposium is to bring together new research looking at the effective use of music in the medical sector and to give impulses for future research in this area.

Neurobiological changes associated with social bonding, stress and flow-state in vocal improvisation

Edward A. Roth[1] and Jason R. Keeler[1]

[1] Brain Research and Interdisciplinary Neurosciences (BRAIN) lab, School of Music, Western Michigan University, Kalamazoo, MI , U.S.A.

People often report a feeling of unity and trust when they perform music together; however, little is known about the neurobiological processes that are associated with that experience. Increased understanding of these neurobiological mechanisms could have clinical implications for people with social deficits in particular. In this feasibility study we explored neurobiological and self-reported measures associated with social bonding, stress, and flow state during group vocal jazz improvisation. Neurobiological measures included the neuropeptide oxytocin (previously shown to regulate social behavior and trust) and adrenocorticotropic hormone (ACTH), which may in part mediate the engagement and arousal effects of music. Flow state is a positive intrapersonal state experienced when the difficulty of a task matches or slightly exceeds one's current ability to perform the task. Flow has been positively correlated with the quality of interpersonal relationships and may be achieved through various music tasks. A within-subjects, repeated measures design was used to assess changes in oxytocin and ACTH during group vocal jazz (quartet) singing in two conditions: pre-composed and improvised. Self-reported experiences of flow were measured after both conditions. The primary aim of this small-sample feasibility study was to evaluate the research protocols. To our knowledge, this study was among the first to assess neurobiological measures in vocal jazz improvisation. Our procedures were implemented as planned and only mild modifications of the protocol are indicated prior to a larger scale endeavor. Though not the primary focus of this study, initial neurobiological and flow state data was generated. Group singing was associated with a decrease in ACTH and a mean increase in oxytocin. Flow state was experienced by all participants in both conditions. Although preliminary, group singing and vocal improvisation may influence social relationships and reduce the biological stress response.

Cochlear Implant-Mediated Perception of Music

Charles J. Limb[1]
[1] Departments of Otolaryngology-Head & Neck Surgery and Neurosurgery, University of California San Francisco School of Medicine, San Francisco, CA, USA

Cochlear implants represent a true modern miracle of healthcare today. The ability to restore hearing in children with profound hearing difficulty was once regarded as science fiction. Today, we routinely provide deaf children with the ability to gain high-level speech perception through cochlear implantation. Although this success has been inspiring on many levels, it has also encouraged both deaf patients and the biomedical community to pursue ever more demanding auditory challenges. The ability to hear music represents the most sophisticated and difficult of these challenges, and is arguably the pinnacle of hearing. Through the use of music as both an inspiring target for which to strive and also as a tool from which to learn, we have learned a great deal about the limitations of current cochlear implant technologies. Pitch represents one of the most fundamental attributes of music, and is also the aspect of music that is the most difficult for CI users. In this presentation, we will discuss the implications of limited pitch perception for CI users, and consider several case studies of CI users that have attempted to pursue musical activities despite these obstacles.

The effects of music interventions during caesarean sections on patients' anxiety levels

Nora K. Schaal[1], Carsten Hagenbeck[2], Julius Gilles[2], Sarah Märthesheimer[1], Wolfram Goertz[3], Oliver T. Wolf[4], Tanja Fehm[2] and Philip Hepp[2]

[1]Department of Experimental Psychology, Heinrich-Heine-University, Düsseldorf, Germany

[2]Clinic for Gynecology and Obstetrics, Heinrich-Heine-University, Düsseldorf, Germany

[3]Musicians' clinic, University Hospital, Düsseldorf, Germany

[4]Department of Cognitive Psychology, Institute of Cognitive Neuroscience, Faculty of Psychology, Ruhr-University Bochum, Germany

Several studies have shown that music interventions before and during surgery can lead to reduced anxiety and pain levels of the patient. The use of this positive effect in the case of caesarean sections could be beneficial for the expectant mother but has not found much attention in the literature yet. The present study investigates the effects of music interventions during the caesarean section on subjective (State Trait Anxiety Inventory, visual analogue scale for anxiety) and objective (saliva and amylase from cortisol) measures of anxiety as well as heart rates perceived before, during and after the caesarean on the day of surgery. The patients are randomly allocated to the experimental group, listening to music in the operating room during the caesarean section, or to the control group, undergoing the procedure without any music intervention. Data from a pilot phase ($N = 45$) revealed significant lower heart rates ($p = .048$) and a trend ($p = .061$) of reduced cortisol levels during the caesarean section in the experimental group compared to the control group. The subjective measures showed descriptively lower anxiety levels in women listening to music. These first results are promising and suggest that listening to music during a caesarean section leads to a reduction of anxiety and stress experienced by the expectant mother. This positive effect could also beneficially influence pain perception and recovery from the operation which are interesting areas for future research. 170 participants have completed the study so far and results from a larger sample ($N \sim 250$) will be presented at the conference.

Music-language dissociation after brain damage: a new look at the famous case of composer Vissarion Shebalin

Meta Weiss[1]
[1]Queensland Conservatorium, Griffith University, Brisbane, Queensland, Australia

The Russian composer Vissarion Shebalin (1902-1963) holds a famous place in music psychology. As documented by Alexander Luria (1965), Shebalin suffered strokes that left him severely aphasic, yet he was still able to compose, and produced works admired by his contemporaries. His case is often taken as evidence that music and language processing can be completely dissociated in the brain. However, the musical structure of his compositions pre- and post-stroke has never been systematically examined. By collecting and analyzing pre- and post-stroke string quartets, the current study finds evidence that Shebalin's post-stroke music showed a marked shifted to a pseudo-atonal style and the use of nonfunctional harmony, as well as a more compact formal construction. These findings are interesting in light of recent research and theorizing in cognitive neuroscience regarding relations between the processing of linguistic and musical syntax (e.g., the 'Shared syntactic integration resource hypothesis', Patel, 2003). The complexity of his case is increased by the historical context of Shebalin's work. By conducting archival research and interviews in Russia, I found that that his post-stroke output also coincided with a more relaxed political environment, which may have encouraged musical experimentation. This makes it difficult to ascribe the change in his compositional style entirely to his strokes, but supports the larger point that Shebalin's music was saliently different before and after his strokes, and that he is thus not a simple example of music-language dissociation after brain damage.

Using a Rhythmic Speech Production Paradigm to Elucidate the Rhythm-Grammar Link: A Pilot Study

Rita Pfeiffer[1,2], Alison Williams[3,2], Carolyn Shivers[4] and Reyna Gordon[5,2]
[1]Department of Hearing & Speech Sciences, Vanderbilt University Medical Center, USA
[2]Program for Music, Mind & Society at Vanderbilt, USA
[3]School of Engineering, Vanderbilt University, USA
[4]Department of Human Development, Virginia Polytechnic Institute and State University, USA
[5]Department of Otolaryngology, Vanderbilt University Medical Center, USA

A growing body of work points to a relationship between rhythm abilities and language skills, including grammar. Individual differences in sensitivity to rhythmic aspects of speech could be a mechanism underlying the association between musical rhythm and grammar skills found in our preliminary work. To assess this, we first modified a rhythmic speech production task and conducted a pilot study to determine whether typically developing six-year-olds would be able to do the following: 1) consistently say a given phrase with the same rhythm across multiple trials; 2) maintain a steady internal pacing; and, 3) synchronize their speech with an isochronous metronome.

We employed a modified version of the speech cycling task (Cummins and Port, 1998), in which participants speak short, metrically regular phrases along with an isochronous metronome. Twelve six-year-old children repeated a set of ten phrases (e.g., "beg for a dime") with two patterns (whole note metronome: one click per cycle, and half-note metronome: two per cycle).

Primary speech beats within each phrase were marked using an algorithm and analyzed as follows: synchrony with the metronome (external phase), synchrony between phrases (period), and synchrony within each phrase (internal phase). Preliminary findings show clear metrical organization of speech beats within each phrase. Both the period and external phase analyses revealed that metronome pattern (i.e., presence of a subdivided beat) significantly influenced children's performance on the task, suggesting the importance of external rhythmic cues in scaffolding children's underlying sychronization. Pilot results show promise for future use of the paradigm to test an overlap of speech rhythm sensitivity, musical rhythm perception, and expressive grammar abilities. Ongoing work entails validating the speech cycling paradigm and enhancing its feasibility for assessment and intervention purposes.

Music syntactic processing related to integration of local and global harmonic structures in musicians and non-musicians

Irán Román[1], Takako Fujioka[1]
[1]Department of Music, Stanford University, Stanford, CA, USA

In listening to Western tonal music, an unexpected out-of-key chord is known to elicit an event-related potential (ERP) component called Early Right Anterior Negativity (ERAN) at right frontal electrodes compared to a standard in-key-chord. However, in more realistic musical pieces, a sense of key can constantly move from a key to another closely-related one. Such movements typically follow the global rule called 'circle of 5ths', which describes the relationship between two keys sharing most chords and scale tones. We recorded EEG from musicians and non-musicians to analyze the evoked potential to an out-of-key chord when preceding local patterns follow the global rule. We examined three conditions: Control (as in previous ERAN studies), Sequential, and Non-Sequential, all of which contained the same out-of-key chord, preceded by different chords. The Sequential condition presented three repetitions of a local pattern including the out-of-key chord, while moving through different keys following the global rule. In contrast, the Non-Sequential condition presented the same local pattern three times without following the global rule, jumping across unrelated keys. During the Sequential condition, musicians showed a left-lateralized ERAN, delayed about 50ms compared to the Control condition. This suggests that the integration of local and global information for successful key motions may require left frontal neural resources, compared to simple processing of an out-of-key chord. Furthermore, a right-frontal positivity with latency around 365ms was found in the Non-Sequential condition, perhaps related to local pattern violation and not to syntax processing. We are currently analyzing non-musician data, and hypothesize that the ERAN in the Sequential condition will be right-lateralized, reflecting only integration of the main key. In the Non-Sequential condition we expect non-musicians to process local pattern violation similarly to musicians.

Musical Ability Predicts Adaptation to Foreign-Accented Speech

L. Robert Slevc,[*1] Paulina Arango,[#2] Mary Dumler[*3]

[*]*Department of Psychology, University of Maryland, College Park, MD, USA*

[#] *Escuela de Psicología, Universidad de Los Andes, Santiago, Chile*

[1]slevc@umd.edu, [2]parango@uandes.cl, [3]marypdumler@gmail.com

ABSTRACT

Musical ability is associated with advantages in speech processing, including better ability to understand speech in noise and higher proficiency in second-language phonology. Musicians also outperform non-musicians when learning non-native speech sound contrasts, such as Mandarin lexical tones. This study investigates whether these advantages in speech sound learning extend to the ability to rapidly adapt to foreign-accented speech. Adult participants' processing of accented speech was assessed via their mouse movement trajectories when selecting between phonetically similar images in response to heavily accented speech tokens. Individual differences in musical ability predicted unique variance in participants' degree of adaptation, even after controlling for perceptual acuity and experience with languages and dialects. These data suggest that musical ability is, in fact, related to the ability to adapt to non-native accented speech and lends further support to the idea that a "good ear" for language is related to musical ability.

I. INTRODUCTION

Learning new languages and communicating with non-native speakers is becoming increasingly important in our multicultural and interconnected world. However, learning and communicating in a second language (L2) can be a major challenge for adult L2 learners, especially when it comes to L2 phonology. That is, many otherwise highly proficient L2 learners maintain a strong foreign accent (e.g., Flege et al., 1999), which can impede ease and flow of communication. There is growing evidence that the ability to successfully perceive and produce L2 phonology is related to musical ability; the question posed here is whether musical ability also predicts the ability to understand heavily accented speech.

A relationship between musical ability and accent adaptation makes sense given a growing body of research showing that musical ability is related to a variety of speech sound processing abilities. These advantages may arise because music places high demands on precise auditory processing, reinforced by music's emotional and attentional draw (Patel, 2014). For example, musicians show advantages in processing linguistic prosody (e.g., Thompson et al., 2004) and in perceiving speech in challenging conditions (e.g., Slater et al., 2015). Musicians also show advantages in the perception, production, and learning of L2 phonology (review: Chobert & Besson, 2013). This has been most thoroughly investigated in terms of pitch-based aspects of L2 phonology, namely lexical tone: Musicians outperform non-musicians in discrimination and imitation of foreign lexical tone contrasts (e.g., Wayland et al., 2010) and in the ability to learn tonal words in a L2 (e.g., Bowles et al., in press).

Musical ability is not only related to the ability to process pitch-based aspects of L2, but also to the perception, production, and learning of non pitch-based aspects of L2 phonology (e.g., Christiner & Reiterer, 2015; Kempe et al., 2015; Parbery-Clark et al., 2012). For example, Slevc and Miyake (2006) found that musical ability in adult Japanese learners of English subjects predicted greater proficiency in perception and production of English (L2) phonology, even when controlling for other factors such as age of L2 exposure, length and amount of L2 experience, and short-term memory (see also Milovanov & Tervaiemi, 2011; Posedel et al., 2011).

There is thus considerable evidence that musical ability is related to effective perception, production, and learning of speech sounds. However, this work has generally focused on proficiency in long-term L2 learners (e.g., individuals immersed in an L2 environment) or on explicit short-term learning or discrimination tasks (both using novel speech sounds and in native language tasks such as speech perception in noise). The current study investigates whether musical ability also predicts the ability to rapidly and implicitly adapt to phonetic variation; specifically, whether it can predict the ability to adapt to foreign-accented speech.

Foreign accented speech is potentially challenging for a listener because adult L2 learners typically produce L2 speech sounds that deviate in systematic ways from the way those speech sounds would be produced by a native speaker. For example, Korean does not have a central vowel equivalent to the English /ɝ/ (as in "work") and so a Korean speaker of English might instead substitute the back vowel /ɔ/ (as in "walk"). Because this likely happens systematically, listeners can potentially adjust their perceptual categorization of phonemes (at least when listening to a given accent or speaker) to learn that something that sounds like "walk" is likely to actually be the word "work". This accent adaptation happens quickly and seemingly automatically (e.g., Bradlow & Bent, 2008; Clarke & Garrett, 2004; Maye et al., 2010; Witteman et al., 2015) however individuals vary in exactly how quickly and effectively they can adapt to foreign accented speech (e.g., Banks et al., 2015; Janse & Adank, 2012).

The present study thus investigated adaptation to heavily accented English over a short (approximately 15 min) task. Because individual differences in accent adaptation are likely to be relatively subtle, we employed a mouse-tracking paradigm (Freeman & Ambady, 2010) where participants' heard accented words (e.g., "bowl") and chose between two pictures corresponding to the intended word (a bowl) and a likely foil (a ball). The extent to which participants mouse movements deviated toward the foil picture (even when correctly choosing the target) was used as a measure of their uncertainty about whether they heard the intended vowel (e.g., /o/) or the vowel in the foil (e.g., /ɔ/).

Individual differences in this accent adaptation task were investigated as a function of musical ability, measured both with a comprehensive questionnaire on musical experience and engagement (Müllensiefen et al., 2014) and a behavioral musical discrimination task (Wallentin et al., 2010). In addition, we assessed a variety of potential confounding factors, including handedness, bilingualism, dialect exposure, and perceptual acuity. If musical ability can predict unique variance in the ability to adapt to a foreign accent even after controlling for these potential confounds, this would suggest that musicians' advantages in speech sound processing extend to the ability to rapidly adjust and adapt to non-standard segmental (non pitch-based) aspects of phonology.

II. METHODS

A. Participants

Eighty-three native-English speaking participants without Korean language experience but with normal hearing (by self-report) were recruited from the University of Maryland, College Park and surrounding area. Fifty of the participants (60.2%) self identified as musicians. Participants either received course credit or $10/hour for their participation.

B. Stimuli and Procedure

1) Accent Adaptation Task. Word and sentence stimuli were selected from the Wildcat Corpus of Native- and Foreign-Accented English (Van Engen et al., 2010) and the Korean-English Intelligibility Database, both via the Online Speech/Corpora Archive and Analysis Resource (OSCAAR). Both databases contain words and sentences recited in English by native Korean speakers. Words and sentences were selected such that, given the speaker's accent, they could easily be interpreted as sounding like either of two distinct words (e.g., "ball" vs. "bowl") both of which could easily be represented with a picture. Pictures corresponding to the target and foil referents of each stimulus were selected and judged for appropriateness and representational aptness by three lab members.

On each trial, participants clicked a "start" button at the bottom of the screen, heard an accented word (or sentence), then moved the mouse to click whichever of two pictures best corresponded to the spoken stimulus (target and foil pictures appeared in the top left or right corner equally often; see Figure 1 for a schematic). Participants first responded to a baseline block of 19 accented words, then to a block of 35 accented sentences, in which contextual information provided an (additional) opportunity to learn how the speaker produces English phonemes, and finally to a post-exposure set of 19 accented words. The task was administered with the MouseTracker software package (Freeman & Ambady, 2010). Dependent measures included accuracy and, for correct trials, response time (RT), the maximum deviation (MD) of the participant's mouse trajectory from a straight-line response to the target, and the area under the curve(s) (AUC) created by plotting the participant's mouse trajectory against a straight-line response to the target.

2) Musical Ability Measures. Musical ability (used broadly here to encompass individual differences in both aptitude and experience) was assessed with a self-report measure and a behavioral discrimination task. The Goldsmith Musical Sophistication Index (Gold-MSI; Müllensiefen et al., 2014) is a set of 39 questions that target multiple aspects of musical experience, engagement, and abilities. These questions load on five distinct components of musical sophistication (active engagement, perceptual abilities, musical training, singing abilities, and emotions) as well as a general sophistication factor. Because these factors are highly related, for the present study we focused on the general sophistication factor.

The Musical Ear Task (Wallentin et al., 2010) requires participants to make same/different judgments on pairs of musical stimuli. In one block (MET-Melody), half of the 52 pairs differ in a single pitch, and half of these pitch differences also change the melodic contour. In the second block (MET-Rhythm), half of 52 pairs of rhythms (played in wood block beats) contain a single rhythmic change. Performance on each subtest was evaluated with discriminability (d-prime) scores.

3) Perceptual Acuity Measures. Perceptual discrimination thresholds were assessed with a set of auditory discrimination tasks, which use an adaptive psychophysical procedure to yield an estimate of the minimum detectable difference between two types of stimuli (from the Maximum Likelihood Procedure Matlab toolbox; Grassi & Soranzo, 2009). The four tasks measured discrimination of pitch, temporal order, gap duration, and intensity.

4) Additional Measures. Participants also completed a questionnaire assessing handedness (Oldfield, 1971), language experience (based on the LEAP-Q; Marian et al., 2007), and their degree of exposure to different dialects. Note that these participants also participated in two additional speech sound learning paradigms, which are not reported here.

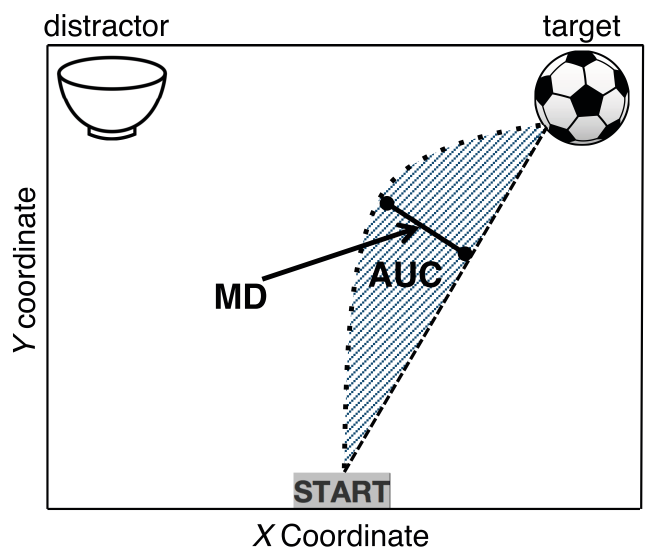


Figure 1. Schematic of the accent adaptation task and mouse trajectory measures. Participants heard an accented word (e.g., "ball") after clicking START, then moved the mouse to click on the appropriate picture. *MD* indicates the maximum deviation of a participant's response from a direct (straight-line) response to the target; *AUC* indicates the area under the curve that separates the participant's response from a direct trajectory to the target.

C. Analysis

Adaptation to foreign accented speech was operationalized as the difference between performance on the baseline and the final post-exposure block of accented words (contrast coded as *block* in the analyses below and shown as a difference score in Figure 2). A set of repeated measures regressions were fit for each of the dependent measures in the accent adaptation task (accuracy, RT, MD, and AUC) as a function of *block* (baseline or post-exposure) and interactions between *block* and musical ability, perceptual acuity, handedness, bilingualism, and dialect exposure.

III. RESULTS

A. Correlations between IVs.

The different factors from the Gold-MSI questionnaire were generally well correlated. We relied on the general sophistication factor, which correlated highly with the other factors (rs between .55 and .89, except for active engagement; $r = .15$). General sophistication also correlated highly with the melody subtest of the MET ($r = .54$, $p < .0001$). However, while the MET subtests were themselves moderately well correlated ($r = .41$, $p < .0001$), the MET rhythm subtest was unrelated to general sophistication ($r = .16$, *n.s.*). Nevertheless, given that these tasks are all theoretically interrelated, we formed a single aggregate measure of musical ability for the analyses below by summing standardized scores on the general sophistication subscale of the Gold-MSI with standardized scores on the two MET subtests. Similarly, the four measures of perceptual acuity were significantly correlated (all ps < .01), thus a single aggregate measure was created for these as well following the same procedure.

B. Regression Analyses.

Surprisingly, accuracy was lower in the post-exposure block compared to the baseline block (79.3% vs. 87.5%), as shown by a main effect of *block* ($b = -0.084$, $t = -8.14$, $p < .0001$). In contrast, response times (RT) were slightly, but significantly, faster in the post-exposure block compared to baseline (2119 ms vs. 2188 ms; a main effect of *block* on log transformed RTs: $b = -0.038$, $t = -2.33$, $p < .05$). Neither the accuracy nor RT effects of *block* differed as a function of individual differences in musical ability or perceptual acuity.

Measures of mouse movement trajectories also differed between blocks. Maximum deviations (MD) showed a main effect of *block* ($b = 0.043$, $t = 2.80$, $p < .01$), which was qualified by an interaction with *musical ability* ($b = -0.017$, $t = -2.38$, $p < .05$). Similarly, area under the curve (AUC) differed across *block* ($b = 0.077$, $t = 2.54$, $p < .05$), and this effect interacted with *musical ability* ($b = -0.029$, $t = -2.05$, $p < .05$). AUC was also slightly lower for participants with lower perceptual discrimination thresholds (a main effect of *perceptual acuity;* $b = 0.022$, $t = 2.14$, $p < .05$), although this main effect did not interact with change across blocks.

As with the accuracy data, these main effects of *block* indicate that mouse trajectories actually showed *greater* deviation from the target in the post-exposure block compared to the initial baseline block. That is, participants' mouse movements appeared to reflect greater uncertainty after they had heard a series of context-rich sentences, and so presumably could have adapted to the accent patterns of the recorded speaker. However, the interaction effects show that, while this was true for participants with lower musical ability scores, participants with higher levels of musical ability showed a (predicted) *reduction* in deviation from targets (i.e., were less influenced by distractors) after exposure. This is illustrated via difference scores in Figure 2, which shows MD scores in the baseline block minus the post-exposure block, after factoring out effects of handedness, bilingualism, perceptual acuity, and degree of dialect exposure.

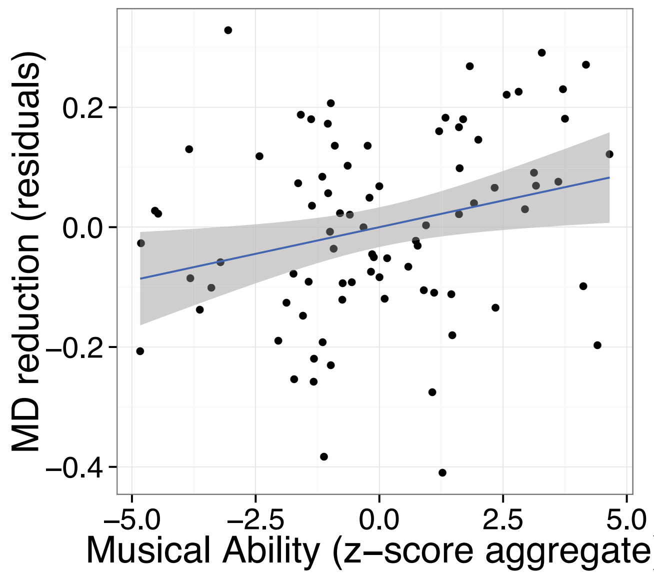


Figure 2. Residual MD (maximum deviation) difference scores (baseline minus post-exposure MD scores), after controlling for handedness, bilingualism, dialect exposure, and perceptual acuity, as a function of performance on the musical ability aggregate.

In contrast to these interactions with musical ability, *block* did not interact with handedness, bilingualism, perceptual discrimination thresholds, or degree of exposure to different dialects for either MD or AUC.

Musical ability predicted higher accuracy ($b = 0.0060$, $t = 2.32$, $p < .05$) and slower response times ($b = 0.017$, $t = 2.11$, $p < .05$) in the intervening block of sentence stimuli. However neither musical ability nor the other IVs were related to the mouse trajectory measures in the sentence block.

IV. DISCUSSION

Musical ability, assessed both by self-report and behavioral discrimination tasks, predicted adaptation to foreign accented speech as assessed by the relative change in mouse trajectory deviations away from a target picture (toward a foil) between a baseline and a post-exposure block of word response trials. This effect emerged even after controlling for individual differences in perceptual acuity, handedness, bilingualism, and exposure to different dialects, suggesting that the relationship between musical ability and accent adaptation does not simply reflect enhanced perceptual skills.

Surprisingly, mouse movement trajectories actually deviated more from a direct trajectory to the target in the post-exposure block than in the initial baseline block overall, although this was true primarily for participants with lower levels of musical ability. A likely explanation for this unexpected pattern is that the words in the post-exposure block were simply more difficult than those in the baseline block, thereby obscuring any signs of improvement. Although words were selected to be of approximately equal difficulty in the two blocks (based on pilot testing), they were not counterbalanced across blocks so that each participant had the same experience, allowing for individual differences analyses.

An alternative possibility is that the overall reduction in performance in the second block does, in fact, reflect a greater 'draw' toward the foil picture after experience, at least for the less musically sophisticated participants. If so, this pattern may reflect a lessening in confidence (reflected by more variable mouse movements) as participants learned that the speaker produces phonemes in an unusual way. For those more musical participants, this may have been balanced out or overcome by learning the systematic ways in which the speaker's accent differs from standard English pronunciation.

Form these data, it is not clear if musically sophisticated participants learned something about the accent generally or about the specific speaker they heard (cf. Bradlow & Bent, 2008). Nor is it possible, from this correlational study, to know the extent to which this musical ability advantage relates to musical *experience*, specific *types* of musical experience, or to musical *aptitude*. While musical training can affect auditory acuity and speech sound learning (e.g., Tierney et al., 2015), and the type of musical experience is likely relevant (e.g., Carey et al., 2015), there are also clearly notable differences between individuals who do and do not pursue musical experiences (e.g., Schellenberg, 2015) and at least some of these pre-existing differences probably also relate to sound category learning abilities.

V. CONCLUSION

These data indicate that individuals with higher levels of musical ability show more robust adaptation to foreign accented speech, even after controlling for other factors including dialect exposure and individual differences in perceptual acuity. This suggests that musical ability is related to the ability to rapidly adjust perceptual categories for speech sounds in order to accommodate systematic deviations from native-like speech that are attributable to a foreign accent. These data complement evidence for other types of speech processing advantages associated with musical ability, including the ability to perceive speech in noise (e.g., Slater et al., 2015) and to successfully learn the phonology of a second language (e.g., Slevc & Miyake, 2006).

ACKNOWLEDGMENT

The work reported here was supported by a grant from the GRAMMY Foundation. We thank Rochelle Newman, Mattson Ogg, and Nabil Morad for assistance and advice.

REFERENCES

Banks, B., Gowen, E., Munro, K.J., & Adank, P. (2015). Cognitive predictors of perceptual adaptation to accented speech. *Journal of the Acoustical Society of America,* 137, 2015-24.

Bradlow, A.R. & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, 106, 707-29.

Carey, D., Rosen, S., Krishnan, S., Pearce, M.T., Shepherd, A., Aydelott, J., & Dick, F. (2015). Generality and specificity in the effects of musical expertise on perception and cognition. *Cognition*, 137, 81-105.

Chobert J. & Besson M. (2013). Musical expertise and second language learning. *Brain Sciences*, 3, 923-940.

Christiner, M. & Reiterer, S.M. (2015). A Mozart is not a Pavarotti: singers outperform instrumentalists on foreign accent imitation. *Frontiers in Human Neuroscience*, 9, 482.

Clarke, C.M. & Garrett, M.F. (2004). Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America,* 116, 3647-58.

Flege, J.E., Yeni-Komshian, G.H., & Liu, S. (1999). Factors affecting degree of foreign accent in an L2: A review. *Journal of Memory and Language*, 41, 78-104.

Freeman, J.B. & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, 42, 226-41.

Grassi, M. & Soranzo, A. (2009). MLP: A MATLAB toolbox for rapid and reliable auditory threshold estimation. *Behavior Research Methods, 41*, 20-8.

Janse, E., & Adank, P. (2012). Predicting foreign-accent adaptation in older adults. *Quarterly Journal of Experimental Psychology*, 65, 1563-85.

Kempe, V., Bublitz, D., & Brooks, P. J. (2015). Musical ability and non-native speech-sound processing are linked through sensitivity to pitch and spectral information. *British Journal of Psychology, 106*(2), 349-66.

Marian, V., Blumenfeld, H.K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in Bilinguals and Multilinguals. *Journal of Speech Language and Hearing Research*, 50, 940-67.

Milovanov, R., & Tervaniemi, M. (2011). The Interplay between Musical and Linguistic Aptitudes: A Review. *Frontiers in Psychology*, 2, 321.

Müllensiefen D., Gingras B., Musil J., Stewart L. (2014). The musicality of non-musicians: an index for assessing musical sophistication in the general population. *PLoS ONE* 9(2): e89642.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9, 97-113.

Parbery-Clark, A., Anderson, S., Hittner, E., & Kraus, N. (2012). Musical experience strengthens the neural representation of sounds important for communication in middle-aged adults. *Frontiers in Aging Neuroscience*, 4, 30.

Patel, A.D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing Research,* 308, 98-108.

Posedel, J., Emery, L., Souza, B., & Fountain, C. (2011). Pitch perception, working memory, and second-language phonological production. *Psychology of Music*, 40(4), 508-517.

Schellenberg, E.G. (2015) Music training and speech perception: a gene-environment interaction. *Annals of the New York Academy of Sciences*, 1337, 170-7.

Slater J., Skoe E., Strait D.L., O'Connell S., Thompson S., Kraus N. (2015). Music training improves speech-in-noise perception: longitudinal evidence from a community-based music program. *Behavioral Brain Research*, 291, 244-252.

Slevc, L.R. & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science*, 17, 675-81.

Thompson, W.F., Schellenberg, E.G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion, 4*, 46-64.

Tierney, A.T., Krizman, J., & Kraus, N. (2015). Music training alters the course of adolescent auditory development. *Proceedings of the National Academy of Sciences,* 112, 10062-7.

Van Engen, K.J., Baese-Berk, M., Baker, R.E., Choi, A., Kim, M., & Bradlow, A.R. (2010). The Wildcat Corpus of Native- and Foreign-Accented English: Communicative Efficiency across Conversational Dyads with Varying Language Alignment Profiles. *Language and Speech*, 53, 510-540.

Wallentin, M., Nielsen, A. H., Friis-Olivarius, M., Vuust, C., & Vuust, P. (2010). The Musical Ear Test, a new reliable test for measuring musical competence. *Learning and Individual Differences*, 20, 188-196.

Wayland R., Herrera E., & Kaan E. (2010). Effects of musical experience and training on pitch contour perception. *Journal of Phonetics*, 38, 654–62.

Witteman, M.J., Bardhan, N.P., Weber, A., & McQueen, J.M. (2015). Automaticity and stability of adaptation to foreign-accented speech. *Language and Speech*, 52, 168-89.

Music Training Correlates with Visual but not Phonological Foreign Language Learning Skills

Peter Martens,[*1] Kimi Nakatsukasa,[#2] Hannah Percival[*3]

[*]School of Music, Texas Tech University, Lubbock, TX, USA

[#]Department of Classical and Modern Languages and Literature, Texas Tech University, Lubbock, TX, USA

[1]peter.martens@ttu.edu, [2]kimi.nakatsukasa@ttu.edu, [3]hannah.percival@ttu.edu

ABSTRACT

Recent studies in music psychology have reported a positive effect of musical training on the development of a child's first language. We hypothesize that a similar effect will exist for foreign language (L2) learning skills. In a pilot study, two subject groups were matched in terms of age and gender; a musician group of advanced students in music degree programs, and a non-musician group of students with little or no musical training. All subjects completed the Language Learning Aptitude Measure (LLAMA) which tests subjects' abilities to learn an artificial language. Results showed small positive effects for the musician group learning novel lexical items by sight, in relating sounds to a writing system, and in inferring visual grammatical rules. There was no positive effect for musicians on a phonological sound recognition task. Two subsequent experiments were conducted with collegiate music majors. In the first, subjects completed LLAMA (n=79), and their results were compared with those of a matched general population from a 2015 study (n=107). Results were strikingly similar to the pilot study, with differences rising to the level of significance. The Advanced Measures of Music Audiation (AMMA) was administered to 23 of the same music majors, yielding scores in tonal and rhythmic musical aptitude. Significant positive correlation between AMMA and LLAMA scores was limited to AMMA rhythmic scores and LLAMA sound recognition scores. Overall these findings suggest that musical training relates positively to visual, aural-visual, and syntactic-visual aspects of L2 learning. The correlation of rhythmic aptitude and phonological learning within musicians, coupled with no difference overall between the musicians and general population on the same task, may suggest that music training can enhance this aspect of L2 learning, but that the general population also has "good ears" for the same information, perhaps stemming from modern habits of music listening.

I. INTRODUCTION

Recent studies in music psychology have reported a positive effect of musical training on the development of a child's first language (e.g. Forgeard et al., 2008). We hypothesized that a similar effect will exist for foreign language (L2) learning skills, especially in aural learning tasks (Slevc & Miyake, 2006). In a pilot study, two groups of subjects were matched in terms of age and gender; the musician group (N=10) consisted of students in music performance majors, while the non-musician group (N=7) consisted of non-music students with little or no musical training. All subjects completed the Language Learning Aptitude Measure (LLAMA; Maera, 2005), which tests subjects' abilities to learn an artificial language; LLAMA consists of four subtests, LLAMA B, D, E, and F. Results showed small positive effects for the musician group learning novel lexical items by sight (LLAMA B), in relating sounds to a writing system (LLAMA E), and on visual grammatical rule inference (LLAMA F). There was no such effect for musicians in implicit learning via sound recognition (LLAMA D). These results were consistent with our overall hypothesis, but differed notably in detail.

II. EXPERIMENT 1

A. Method

To investigate these findings further, we conducted a full study using college music major subjects, all of whom had had at least three semesters of training in theory, aural skills, and in their primary performance area (n=79). Music major results were compared with a general population of native English speakers (n=107), tested with the LLAMA battery and reported by Rogers et al. (2015).

Our primary initial task was to determine that these subject populations were adequately matched. One of the main purposes of Rogers et al. (2015) was to identify subject factors that were and were not predictive of LLAMA performance, which allowed us to conclude that the populations were sufficiently similar to be compared. Age, gender, and educational level were found to account for only a small amount of the performance variability on any of the four LLAMA tests, and will not be discussed further. Some factors of language background and current L2 study were found to correlate with LLAMA performance, the most important of which was primary language. Thus only L1 English speakers from the Rogers et al. study were used in this comparison, matching the music major subject pool of exclusively of English L1 speakers. While some music major subjects were bilinguals, Rogers et al. (2015) found no significant differences between mono- and bilinguals on any LLAMA test. Finally, over half of subjects tested in Rogers et al. (2015) were current L2 learners, who were found to outperform both mono- and bilinguals on LLAMA B and F. None of the music major subjects were L2 learners; thus, we might expect that any advantage of music training on those two tests would be counterbalanced by the L2 advantage of the Rogers et al. (2015) subjects.

B. Results

Despite the presence of L2 learners in this general population, results of Experiment 1 closely mimicked those of the pilot study. Music majors outperformed Rogers et al. (2015) subjects on LLAMA B, E, and F, with no significant difference found between LLAMA D scores, as shown in Figure 1. A two-tailed Welch's *t*-test was used to account for the unequal number of subjects and to avoid assuming equal variance, with these comparisons yielding the following values: LLAMA B ($p<.0001$, $d=.652$); LLAMA D (not sig.); LLAMA E ($p=0.0004$, $d=.453$); LLAMA F ($p<.0001$, $d=.627$).

ISBN 1-876346-65-5 © ICMPC14

have contributed to their superior performance on LLAMA B, E, and F.

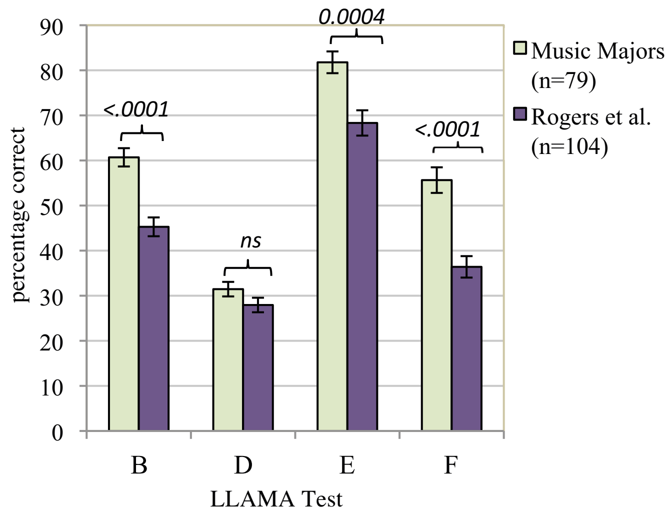


Figure 1. Experiment 1 results, showing percentage correct on LLAMA tests by Music Majors and Rogers et al. (2015) subjects (*p*-values from two-tailed Welch's *t*-test).

C. Discussion

These results suggest to us that the similar performance between musicians and non-musicians on LLAMA D in the pilot study (phonological recognition) were not, as originally thought, anomalous or due to small sample size. Two subject factors seem important in considering these results. First, members of a culture are generally expert music listeners in that culture via extensive exposure (Sloboda, 2011; Margulis, 2014). This is especially true for the youth of first-world cultures, given the ubiquity of recorded music and personal access to it. Success in LLAMA D depends solely on short-term recall of unfamiliar phonemes, and of unknown words made up of both familiar and unfamiliar phonemes. Might it be that the explicit and implicit aural training of music majors does not enhance their skills beyond the already high skill level granted implicitly to their (ostensibly untrained) peers in the general population?

Another factor to note involves the nature of music pedagogy and modes of music-making in the western world. Notational literacy is prioritized, even for beginners, in school-based instrumental and choral music programs. Even aurally-focused early instruction such as the Suzuki Method typically incorporates note-reading during the early or middle elementary school years. The ensemble orientation of high school music programs depends in large part on literacy and on an increasingly sophisticated grasp of notational and expressive nuance on the printed page. Further, college music majors spend many hours per week engaged in score reading and analysis as part of instruction in the private studio and music theory classes, with the goal of connecting "sound to symbol" and vice versa frequently made explicit in these contexts. The above results suggest that music majors' enhanced abilities to interpret non-linguistic symbols, to quickly link and remember written words or sounds that correspond to those symbols, and to infer syntactic relationships between unfamiliar words and symbols may have contributed to their superior performance on LLAMA B, E, and F.

III. EXPERIMENT 2

A. Method

If the superior performance by music majors observed in Experiment 1 were indeed attributable to their training and experience in music, then the degree to which these music majors demonstrate proficiency in basic musical skills might also be reflected in their individual LLAMA scores. 23 music majors who participated in Experiment 1 also took the Advanced Measures of Musical Audiation battery (AMMA; Gordon, 1989), in which they listened to matched-pair melodies of 15 to 20 notes each. These pairs were either identical, they differed in one or more pitches (with rhythms identical), or they differed in one or more rhythmic durations (with pitches identical). Subjects had three response choices: Same, Different-Tonal, or Different-Rhythm. Each subject received a "Tonal" score for detecting pitch-based differences (and the absence thereof) between melodic pairs, and a "Rhythm" score for likewise detecting rhythm-based differences (and the absence thereof).

B. Results

AMMA scores (Tonal and Rhythm) and LLAMA scores (B, D, E, and F) were paired within subjects and analyzed via Pearson's correlation (df=21). The only comparison that returned a significant result was AMMA Rhythm and LLAMA B (r=.386, one-tailed p=.0345)

C. Discussion

If we consider what musical skills the AMMA battery is testing, these results are not surprising in light of the discussion following Experiment 1. While trained musicians have been shown to outperform non-musicians on the AMMA battery (Gordon, 1989), the perceptual acuity in pitch and rhythm on which LLAMA D performance depends is not nearly so fine-grained as needed for AMMA. Thus, while music majors' AMMA scores were predictive of their LLAMA D scores, music majors as a group did not outperform a general population on LLAMA D. At the same time, the aural/visual/logical skills used for LLAMA B, E, and F were not measured by AMMA, and thus the lack of correlation is not surprising.

IV. GENERAL DISCUSSION

Given the links suggested above between literate musicianship and visual-based L2 learning skills, further work is needed with expert musicians trained in an oral rather than a written tradition. While a completely notation-illiterate group of musicians may be difficult to identify within the L1 English speaking population, such work would shed further light on whether or not the specific type or method of music training determines the nature its correlation with LLAMA scores. On the L2 aptitude side of the equation, it may be that music training is positively correlated with some phonological abilities (e.g. perception of discriminatory phonemes, as in Slevc & Miyake, 2006) and not others (e.g. sound recognition, as here). Finally, the perhaps surprising visual and/or intellectual basis for the correlations above may play into the association of general intelligence with music aptitude and training (Schellenberg, 2012). The fundamental musical aptitudes or skills typically assessed within performance-focused music programs (i.e. pitch and rhythm perception and execution) may not represent the sole or even primary enhancements to learning (L2 or otherwise) that correlate with music training.

ACKNOWLEDGMENT

The authors wish to thank David Villarreal and his staff in the Texas Tech University Language Learning Lab for their logistical assistance with data-gathering.

REFERENCES

Forgeard, M., Winner, E., Norton, A., Schlaug, G. (2008). Practicing a Musical Instrument in Childhood is Associated with Enhanced Verbal Ability and Nonverbal Reasoning. *PLoS ONE 3(10):* e3566.

Gordon, E. E. (1989). *Advanced Measures of Music Audiation (Grade 7 – Adult).* Chicago, IL: GIA Publications.

Margulis, E. H. (2014). *On repeat: How music plays the mind.* Oxford: Oxford University Press.

Meara, P. M. (2005). *Llama language aptitude test*. Swansea, UK: Lognostics.

Schellenberg, E. G. & Weiss, M. W. (2012). Music and Cognitive Abilities. In Deutsch, D. (Ed.), *The Psychology of Music* (3rd ed.), pp. 499-550. Cambridge, MA: Academic Press.

Slevc., L. R. & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science, 17*, 675-681.

Sloboda, J. (2011). Musical Expertise. In Levitin, D. J. (Ed.), *Foundations of Cognitive Psychology: Core Readings* (2nd ed.). Boston: Allyn & Bacon.

Rogers, V., Barnett-Legh, T., Curry, C., Davie, E., & Meara, P. (2015). Validating the LLAMA aptitude tests. Presented to the European Second Language Association (EUROSLA25), August 27-29, 2015, Aix-en-Provence, France.

Musical Rhythm in Typical and Atypical Language Development

Organizers: J. Devin McAuley[1,2] and Reyna L. Gordon[3,4]
Discussant: Sarah Creel[5]
[1]Psychology, Michigan State University, East Lansing, MI, USA
[2]Neuroscience, Michigan State University, East Lansing, MI, USA
[3]Otolaryngology, Vanderbilt University Medical Center, Nashville, TN, USA
[4]Program for Music, Mind & Society at Vanderbilt, Nashville, TN, USA
[5]Cognitive Science, UC San Diego, La Jolla, CA, USA

The importance of rhythmic ability for normal development of language skills, in speech and in print, is increasingly well-documented. The broad range of neurodevelopmental disorders that implicate rhythm and/or temporal processing as a core deficit is also quite striking. These include developmental stuttering, developmental dyslexia, and specific language impairment (SLI), among others. This symposium brings together four scholars working at the intersection of musical rhythm and language to consider (a) the role of rhythm in typical language development and (b) the specific nature of rhythm and/or temporal processing deficits in different language disorders.

Behavioral and neuroimaging evidence for a core deficit in rhythm processing in developmental stuttering

J. Devin McAuley[1,4], Liz Wieland[2], Ho Ming Chow[3], Juli Wade[1,4], and Soo-Eun Chang[3]
[1]Psychology, Michigan State University, East Lansing, MI, USA
[2]Communicative Sciences and Disorders, Michigan State University, East Lansing, MI, USA
[3]Psychiatry, Univesity of Michigan, Ann Arbor, MI, USA
[4]Neuroscience, Michigan State University, East Lansing, MI, USA

Stuttering is a neurodevelopmental disorder, characterized by frequent occurrences of involuntary disfluent speech patterns, that affects approximately 1% of the population. Notably, when speech is synchronized with an external rhythmic pacing signal (e.g., a metronome), stuttering can be markedly alleviated even in the most severe cases. These findings suggest that people who stutter may have a core deficit in the generation of an internal rhythm to guide the pacing of speech that is alleviated with external rhythmic support. For the past several years, our group has investigated this hypothesis through a series of behavioral and neuroimaging studies of children and adults who stutter that have been combined with studies in songbirds, which under controlled conditions can be induced to stutter. Behavioral studies with children and adults have leveraged a rhythm discrimination paradigm in which participants listen to two presentations of a standard rhythm followed by a third (comparison) rhythm that is the same or different from the twice presented standard. Rhythms have an explicit accent on each of the beats (simple rhythms) or do not (complex rhythms). Results have revealed (1) poorer rhythm discrimination in children and adults who stutter compared with age-matched controls, especially for complex rhythms that lack explicit markers of the beat, (2) attenuated functional connectivity during resting conditions within a basal ganglia-thalamic-cortical network previously implicated in rhythm perception, and (3) a robust positive correlation between rhythm discrimination performance and functional connectivity in controls that is absent in people who stutter. Findings provide strong support for a core neural deficit in internal rhythm generation in people who stutter. Relation to on-going studies from our group with songbirds and implications for the development of novel rhythm-based interventions will be discussed.

Musical rhythms, speech rhythms, and brain rhythms: new applications for child language development

Reyna L. Gordon[1,2], Rita Pfeiffer[3,2], Alison Williams[4,2], Katherine B. Jones[3,2], Ashley B. Hirsch[1,2], Tuuli H. Morrill[5], J. Devin McAuley[6]
[1]Otolaryngology, Vanderbilt University Medical Center, Nashville, TN, USA
[2]Program for Music, Mind & Society at Vanderbilt, Nashville, TN, USA
[3]Hearing & Speech Sciences, Vanderbilt University Medical Center, Nashville, TN, USA
[4]Biomedical Engineering, Vanderbilt University, Nashville, TN, USA
[5]English, George Mason University, Fairfax, VA, USA
[6]Psychology, Michigan State University, East Lansing, MI, USA

Musical rhythm and linguistic grammar are both hierarchically organized, and the timing of the speech signal contains important cues to grammar, such as pauses at clause boundaries. Here we present a series of experiments aimed at investigating the role of rhythm in grammatical development, in children with normal and disordered language. Our preliminary work in n=25 typically developing (TD) six-year-olds found a robust association (r=0.70) between music rhythm perception and grammar skills. Performance on rhythm perception tasks accounted for individual differences in children's production of complex sentence structures (complex syntax: r=0.46; and transformation/reorganization: r=0.42). Preliminary findings of EEG patterns (recorded during listening to simple rhythms) indicate that higher grammar scorers show enhanced neural sensitivity to accented tones (p=0.02) at early latencies. Pilot behavioral testing in n=3 children with language impairment (SLI) also show lower musical rhythm scores compared to their TD peers (d=-0.88). To examine mechanisms underlying these associations, we are investigating speech rhythm as a possible mediator between musical rhythm and grammar. First, we are probing children's ability to synchronize speech with an isochronous metronome, in a modified speech cycling paradigm. Preliminary results point to correlations (r=0.60) with grammar skills. Second, we are testing the ability of children with SLI and TD to learn novel word-forms from syllable streams that are rhythmically biased to facilitate particular groupings. Findings will be discussed in the context of our ongoing work exploring the dynamics of rhythm deficits in SLI and studying whether facilitations of rhythm on grammar are due to auditory working memory, speech rhythm sensitivity, or both. This line of research contributes to knowledge about the role of rhythm in acquisition and processing of syntactic structure.

Cognitive Links Between Rhythm Perception and Language: A Behavioral and Neuroimaging Investigation

Ola Ozernov-Palchik[1,3,4], Elizabeth S. Norton[2], Sara D. Beach[3], Anne Park[3], Maryanne Wolf[1], John D. E. Gabrieli[3], Nadine Gaab[4], and Aniruddh D. Patel[5]
[1]Child Study and Human Development, Tufts, Medford, MA, USA
[2]Communication Sciences and Disorders, Northwestern University, Evanston, IL, USA
[3]Brain and Cognitive Sciences, MIT, Cambridge, MA, USA
[4]Developmental Medicine Research, Boston Children's Hospital, Boston, MA, USA
[5]Psychology, Tufts, Medford, MA, USA

Studies have reported a link between phonological awareness and musical rhythmic abilities. Both speech perception and musical rhythm perception involve sensory encoding of timing/intensity patterns, short term memory of temporal patterns, and perception of grouping/phrasing. Beat processing is an important component of musical rhythm, but since speech is not periodic, the association between the processes is not clear. This study investigated the relationship among phonological, musical rhythm, and beat processing, both behaviorally and its neural correlates. As part of a longitudinal functional neuroimaging study of reading ability, children (N=35) performed a phonological (first-sound matching contrasted with voice matching) in-scanner task in kindergarten. In 2nd grade these children were administered a rhythm discrimination task with strong beat (SB) and weak beat (WB) patterns, as well as cognitive and pre-literacy measures. Behavioral results indicated a significant positive association between SB (but not WB) performance and children's phonological awareness and reading skills ($p<0.001$). Mediation analysis revealed a significant indirect effect of rhythm discrimination on literacy skills, $b=.59$, BCa CI [.118, 1.39]. Hierarchical regression analysis demonstrated that temporally regular patterns, but not irregular patterns, accounted for unique variance in reading ($p<0.01$). Children's whole-brain functional activation for first sound matching versus voice matching contrast was significantly ($p<0.01$, corrected) correlated with their subsequent performance on SB and WB patterns. The pattern of correlation differed between SB and WB patterns. Implications of these results will be discussed.

Individual differences and reading skills: another PATH needed?

Nori Jacoby[1], Naftali Tishby[2], Hilla Jakoby[2,3], Merav Ahissar[2]
[1]Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge MA, USA
[2]The Edmond & Lily Safra Center for Brain Sciences,Hebrew University, Jerusalem, Israel
[3]Gonda Multidisciplinary Brain Research Center, Bar Ilan University, Ramat-Gan, Israel

Musicians have better performance in linguistic tasks such as reading abilities (e.g., Hurwitz et al., 1975) and speech in noise (e.g., Parbery-Clark et al., 2009). Recently, Tierney and Kraus (2014) suggested that precise auditory timing is a core mechanism underlying enhanced phonological abilities in musicians. Consistent with this hypothesis, Tierney and Kraus (2013) found a correlation between tapping performance and reading skills also in the general population. We empirically tested the precise auditory timing hypothesis (PATH) on three groups of subjects. In the first study we evaluated the entrainment, reading skills and other cognitive abilities of a group of (N=56) normal adults. In the second, we examined the difference between dyslexic (N=26) and non-dyslexic adults (N=19). In the third, we looked at the contrast between dyslexic (N=26) and non-dyslexic professional musicians (N=42). In the first study we replicated Tierney and Kraus's results, finding correlations between reading skills and entrainment in the normal population (word reading r=-0.4 p=0.034; non word reading r=-.46 p=0.013). However, upon measuring a larger set of additional cognitive measures, we found that these correlations could be explained by a more prominent correlation between entrainment and working memory (WM), peaking in a non-auditory visual N-back task (r=-0.52 p=0.0059). In the second study we found no difference in the entrainment abilities of dyslexic and non-dyslexic adults. In the third study, only dyslexic musicians displayed a correlation to WM abilities (r=.45 p=0.02). Our finding suggests an alternative explanation for the PATH hypothesis. We propose that more general cognitive mechanisms strongly associated with WM mediate the link between reading and entrainment. This interpretation is also consistent with recent computational models of reading disabilities (Sagi et al. 2015).

Individuals with congenital amusia show impaired talker adaptation in speech perception

Fang Liu[1], Yanjun Yin[2], Alice H. D. Chan[3], Virginia Yip[2] and Patrick C. M. Wong[2,4]
[1] Department of Language and Linguistics, University of Essex, Colchester, Essex, UK
[2] Department of Linguistics and Modern Languages, The Chinese University of Hong Kong, Hong Kong SAR, China
[3] Division of Linguistics and Multilingual Studies, School of Humanities and Social Sciences, Nanyang Technological University, Singapore
[4] Brain and Mind Institute, The Chinese University of Hong Kong, Hong Kong SAR, China

Talker adaptation is a mechanism listeners use to accommodate speaker variability in speech perception. In a classical example, listeners categorized a tone as “high” when following a low-pitched context, but “low” in a high-pitched context. This study examined whether individuals with congenital amusia, a disorder of pitch processing, would show talker adaptation for lexical tone perception (Experiment 1), and whether talker adaptation is reflected at the brainstem level (Experiment 2).

Thirteen Mandarin-speaking amusics and 13 controls completed a lexical tone categorization task in Experiment 1, with either no context, or a high- or low-pitched context. The target stimuli were a continuum of Mandarin syllables /yi/, /wu/, and /yu/, with the initial fundamental frequency (F0) between 165-200 Hz and the final F0 at 200 Hz. In Experiment 2, fifteen Mandarin speakers passively listened to /yi/ with initial F0s at 180 or 200 Hz in high- or low-pitched contexts while their frequency-following responses (FFRs) were recorded, in order to determine whether talker adaptation was a high-order cognitive or low-order brainstem process.

For tones with no context, amusics exhibited a shallower slope (i.e., reduced categorical perception) than controls when categorization rates were plotted as a function of initial F0. While controls’ lexical tone categorization demonstrated a significant context effect due to talker adaptation, amusics showed similar categorization patterns across both contexts. However, normal listeners’ FFRs to tones in different contexts did not show a talker adaptation effect.

These findings suggest that congenital amusia impacts the extraction of speaker-specific information in speech perception. Given that talker adaptation appears to be a high-order cognitive process above the brainstem level, the deficits in amusia may originate from the cortex.

Music Supports the Processing of Song Lyrics and Changes their Contents: Effects of Melody, Silences and Accompaniment

Yke Schotanus[1]
[1]ICON, Universiteit Utrecht, Utrecht, The Netherlands

A growing body of evidence indicates that singing might affect the perception of language both in a supportive and in a detrimental way. Accompaniment might be a factor of interest, because loud rests (on-beat silences) and out-of-key notes may confuse listeners, while accompaniment might elucidate rhythm and harmony and reduces the amount of silences. This might further music processing and concentration. The main aim of the experiment was to test this hypothesis in a class room situation and investigate at the same time, whether school children might benefit from the use of song, even after a single exposure. A total of 274 pupils (average age 15, SD 0,6), spread over twelve fixed groups, listened to four complete songs in six different conditions (spoken or sung, with and without accompaniment, sung with 'lala' instead of lyrics, and accompaniment only). In a regular lesson Dutch Language and Literature each group listened to five different tracks and completed a questionnaire after each track. The questions concerned processing fluency, valence, comprehension, recall, emotion and repetition. Listeners rated accompanied songs more beautiful and less exhaustive than unaccompanied songs or spoken lyrics, and had the feeling that in these songs the text is more intelligible and comprehensive. With accompaniment the singers voice was considered more relaxed and in tune than without (although the same recordings were used). Furthermore, the silences in unaccompanied versions were rated more distracting than the interludes or the accompaniment in the other conditions. Lyrics were considered more happy, funny, sensitive and energetic, and less sad, heavy and nagging in the sung versions. Finally, verbatim repetition of words is less accepted when spoken, and adds (mainly) emotional meaning to those words in the sung conditions. These findings indicate that accompanied singing supports the transfer of verbal information, although it affects the contents of the words.

Music is More Resistant to Neurological Disease than Language: Insights from Focal Epilepsy

Laura J. Bird[1,2], Graeme D. Jackson[2,3], and Sarah J. Wilson[1,2]
[1]Melbourne School of Psychological Sciences, The University of Melbourne, Melbourne, Victoria, Australia
[2]The Florey Institute of Neuroscience and Mental Health, Austin Health Campus, Melbourne, Victoria, Australia
[3]Department of Medicine, The University of Melbourne, Melbourne, Victoria, Australia

Language and music are higher cortical functions supporting verbal and non-verbal human communication. Despite their shared basis in auditory processing, music appears more resistant to neurological disease than language. Focal epilepsy provides a good model to test this hypothesis, selectively disrupting unilateral network functions. Language may also reorganise due to early onset left focal seizures, while effects on music functions are less clear. We compared the effects of focal epilepsy on music and language functions relative to age of seizure onset, laterality and seizure network. We expected music to be more resistant to left or right frontal or temporal network seizures than language, reflecting a greater bilateral cortical representation. We tested 73 pre-surgical epilepsy patients and 31 matched controls on a comprehensive battery of music tests (pitch, melody, rhythm, meter, pitch and melodic memory, melodic learning), language tests (verbal fluency, learning, memory, verbal reasoning, naming), visuospatial and executive functions, and IQ. Controlling for age and IQ, patients performed worse than controls on verbal fluency, verbal reasoning, and word list recall (all *ps* $<.05$), associative learning of semantically-linked (p =.065) and arbitrary-linked word pairs (p $<.01$) and their delayed recall (p $<.01$). This was particularly evident for left frontal or temporal seizure foci. A specific rhythm deficit was seen in left foci patients (p $<.05$), but no other differences in music skills between patients and controls. There were no effects for age of seizure onset (all *p*s $>.05$). So while disruption of left-sided networks impairs language, verbal learning and memory, music skills were largely preserved for left or right-sided network disease, irrespective of seizure onset. This supports theories of proximal but distinct neural systems for language and music, with left specialisation for temporal processing, and with music benefitting from greater bilaterality.

An fMRI study comparing rhythm perception in adults who do and do not stutter

Elizabeth A. Wieland[1], J. Devin McAuley[2], Ho Ming Chow[3], David Zhu[4], Soo-Eun Chang[3]
[1]Department of Communication Sciences and Disorders, Michigan State University, USA
[2]Department of Psychology and Neuroscience Program, Michigan State University, USA
[3]Department of Psychiatry, University of Michigan, USA
[4]Department of Radiology and Psychology, Michigan State University, USA

Stuttering is a disorder that affects the timing and rhythmic flow of speech; however, when speech is synchronized with an external pacing signal, stuttering can be markedly alleviated. Several neuroimaging studies have pointed to deficiencies in connectivity of the basal ganglia thalamocortical (BGTC) network in people who stutter, suggesting that stuttering may be associated with deficient auditory-motor integration, timing, and rhythm processing. This study compared brain activity patterns in adults who stutter (AWS) and matched controls performing an auditory rhythm discrimination task while undergoing functional magnetic resonance imaging (fMRI; $n = 18$ for each group). We hypothesized that AWS would show (1) worse rhythm discrimination performance, particularly for complex rhythms, and (2) aberrant brain activity patterns in the BGTC network during the rhythm discrimination task, particularly for complex rhythms. Inside the scanner participants heard two presentations of a standard rhythm and judged whether a third comparison rhythm was the same or different. Rhythms either had an explicit periodic accent (simple) or did not (complex). As expected, simple rhythms were easier to discriminate than complex rhythms for both groups ($p < .001$). Although no main group difference was found, a significant rhythm type by group interaction exists ($p = .038$), where AWS had greater difficulty discriminating complex rhythms than controls. The fMRI results demonstrated that AWS had greater (and more right-lateralized) activation than controls in the BGTC network (e.g., right STG, right insula). Our results suggest that AWS are utilizing the rhythm processing network more during the task than controls, and may mean that AWS have to "work harder" during the task or are less familiar with utilizing this network. This difference may underlie deficiencies in internal rhythm generation, in turn affecting the timing of speech movements that result in disfluent speech.

Cross-domain pitch perception, the influence of native language

Ao Chen[1], M. Paula Roncaglia-Denissen[2,3], Drikus A. Roor[4] and Makiko Sadakata[2,4,5]
[1]Utrecht Institute of Linguistics, Utrecht University, Utrecht, the Netherlands
[2]Institute for Logic, Language and Computation, University of Amsterdam, Amsterdam, The Netherlands
[3]Amsterdam Brain Cognition (ABC), University of Amsterdam, Amsterdam, The Netherlands
[4]Musicology Department, University of Amsterdam, Amsterdam, The Netherlands
[5]Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, The Netherlands

Language and music are both high-level cognitive domains that highly rely on prosody. While whether these two domains correlate has received much attention, the issue of whether experience may shape such correlation is under-investigated. To fill in the void, the current study explores pitch perception in both domains among listeners speaking different languages, and we ask the question whether language experience may influence the correlation between perception of pitch in speech and music. We tested Chinese-English, Dutch-English, and Turkish-English non-musicians bilinguals on their sensitivity to pitch information in speech and music, with their English proficiency controlled for. Chinese is a syllable timed tone language, whereas Dutch and Turkish are stress timed, with Dutch have a predominantly strong-weak metrical structure and Turkish the opposite. For the speech domain, the listeners were tested on their discrimination of Chinese lexical tones. For the music domain, they were tested with the melody part of Musical Ear Test (MET), where they needed to discriminate between two melodic musical phrases. We found that no significant difference between the three groups for the lexical tone discrimination while, for the MET, the Chinese-English and Turkish-English listeners outperformed the Dutch-English listeners. Importantly, for both the Dutch-English and the Turkish-English listeners, significant correlation was found between the lexical tone discrimination and the MET, whereas no such correlation was found among the Chinese listeners. Our results suggest that 1) knowing two prosodically different languages enhances musical pitch perception and 2) Unlike non-tone language listeners who perceive both lexical tones and musical pitch psycho-acoustically, tone language listeners perceive their lexical tones phonemically, and they seem to "split" lexical tones from other non-phonemic pitch variations, resulting in the lack of cross-domain correlation.

Analyzing Linguistic Intonation Using The Implication-Realization Model

Alfred W. Cramer

***Abstract*—**Narmour's Implication-Realization model of musical melody (IR) is a promising framework for the study of linguistic intonation. This presentation demonstrates a computer-assisted method for applying IR to prosody, clarifies the relative significance of pitch and other parameters in intonation as compared to music, and develops IR by exploring its explanatory potential. Analyses involving various languages show that IR is able to account for intonational phenomena including the intersecting use of pitch, duration, and other parameters to mark prosodic features; the operation of intonational features on levels from syllable to phrase to discourse; and the intonational conveyance of information structure.

***Keywords*—**melody, linguistic intonation, prosody, Implication-Realization, Autosegmental-Metrical

I. Introduction

Within music theory, Narmour's Implication-Realization model of melody (IR) [1]–[3] is unusual in that it treats a musical melody as a series of pitched tones not fundamentally perceived in terms of harmony or discrete pitch identity. IR would thus seem to be readily applicable to speech intonation, a type of melody matching the same description. IR holds promise for the study of intonational phonology because it analyzes melody in terms of a system of perceptual binary contrasts, an essential element of phonological description. A crucial element of prosody is its use of accents to convey informational coherence between elements of discourse, and here yet again IR holds promise, if it can be shown that IR's network of implications and realizations within a melody is analogous to intonation's cohesive discursive relationships.

In this paper, I outline the possible use of IR to describe linguistic intonation partly to suggest what IR may offer to linguistics and partly to deepen our understanding of music. IR is especially well suited to comparison with the Autosegmental-Metrical model of intonation (AM) [4]–[8]. AM's contributions include its interpretation of the spoken pitch contour as a series of discrete accents or points of attention; its conception of accent as a phenomenon realizable in features other than pitch, such as intensity or duration; and its relationship to discourse and information structure.

A. W. Cramer is a member of the Music Department and an affiliate of the Linguistics and Cognitive Science Department, Pomona College, Claremont, CA 91711 USA (phone: 909-607-2455; e-mail: acramer@pomona.edu).

II. Adaptation of the Implication-Realization Model

IR is best known for positing that listeners expect an initial small interval (S) to be followed by another small melodic interval in the same direction (a structure denoted P^0 or $P^\sim$) and an initial large interval (L) to be followed by a smaller interval in the opposite direction (R^-), as occurs in Figs. 1 and 2. No less central to the model, however, are its claims about melodic closure, hierarchy, and categorical distinctions between interval sizes. The present paper employs an adaptation of IR intended to capitalize on IR's ability to classify interval sizes based on binary distinctions and to make its calculations of implicativeness and closure easier to follow. While it is in some ways a simplification, the adaptation is intended as an orthographic restatement, not a revision of the substance of Narmour's IR. The adaptation allows the description of pitch as a series of melodic intervals of varied sizes, leading to possibilities not available in linguistic approaches that rest on the classification of discrete pitch levels.

IR models melodic hearing as a process of successively comparing the intervals formed by consecutive pairs of tones. From this series of comparisons a network of both binary and

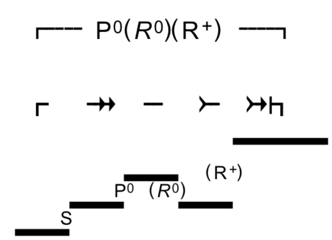


Fig. 1. Illustration of IR analysis. Bracketed structural label as used in [1], [2] at top; this article's labeling system in middle and at bottom.

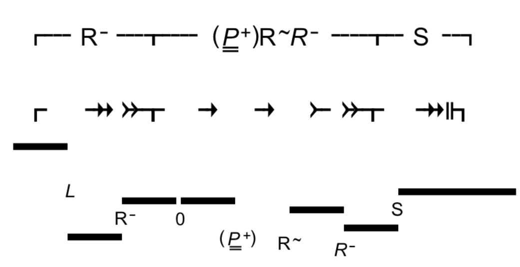


Fig. 2. A second illustration of IR analysis.

TABLE 1
IMPLICATION-REALIZATION SYMBOLS

Interval type	Feature described	Implication/closure symbol
0	initial interval of zero magnitude	→
P	Process (continuation of previous interval's direction)	–
R	Reversal of previous direction	⤚
D	Second consecutive interval of zero magnitude	–
+	Interval contrastively larger than previous	→
-	Interval contrastively smaller than previous	⤚
0	Interval precisely same size as previous	–
~	Interval approximately same size as previous	–
P̲	Non-zero interval following zero interval	–
R̲	Zero-magnitude interval after non-zero interval	⤚
	Full closure	¬
	Implication suppressed (closure promoted) by feature other than interval or pitch	ǀ or –ǀ
	Closure suppressed by non-pitch feature	→
( ) [parentheses]	Retrospective (unexpected) realization	
No parentheses	Prospectively expected interval	
	Initial tone	⌐
S	small initial interval	↠
L	large initial interval	↠
0	Initial interval of magnitude 0	→
Non-Italic letter	Ascending interval	
Italic letter	Descending interval (somewhat closural)	

scaled contrasts emerges in which pitch relationships are specified fairly precisely. The tones express both long-term and short-term implicative connections between tones, and these connections can frame events as either expected or surprising. Given two consecutive intervals (spanning three tones), IR in effect describes the later interval by comparing it to the earlier interval, using interval-size labels that combine the symbols P, R, and D with the symbols $^+$, $^-$, $^\sim$, and ^{0} as shown in Table 1. A reversal in direction R or a sufficient decrease in size $^-$ from the earlier interval contributes closure, and any such closural feature is indicated by an arrowtail ⤚ above the final tone. Conversely, a sufficient increase in interval size is implicative (anticlosural) and is indicated by an arrowhead →. (This usage of arrowheads and tails departs somewhat from Narmour's, where arrowheads and tails typically describe implicative connections. Here they describe changes in what Margulis calls expectancy [9].) Non-pitch features that contribute closure (for example, a clear increase of a tone's duration over that of the previous tone) are indicated by vertical bars ǀ. Features that suppress closure (such as a tone's clearly decreased duration with respect to the previous tone) are indicated by a small, high arrow $^{\rightarrow}$. A tone's net implication is:

$$\text{net implication} = (i - c) + (s_c - s_i) \qquad (1)$$

where i = arrowheads and c = tails occurring since the beginning of the structure, and where and s_c = small arrows and s_i = vertical bars applying to the tone itself.

If net implication > 0, the listener proceeds to the next tone, now treating the old "later" interval as the new "earlier" interval against which to compare the new "later" interval formed by the new tone. If net implication ≤ 0, the tone is fully closural ¬ and is expressed on the next hierarchical level, where it interacts with the earlier and later tones that project to that level.

The initial interval of a segment—including any interval following a break or begun by a closural tone—is not evaluated in comparison to the previous interval. Depending on its size. Its type is labeled L or S, its initial tone is marked ⌐, and its second tone normally receives two arrowheads ↠ (see Table 1).

For comparison, Narmour does not use the symbols L and S, and his letter symbols refer to three-note structures, not to two-note intervals. Analyses are often presented in a kind of shorthand involving brackets that explicitly denote only the hierarchically salient beginnings and endings of structures (see the tops of Figs. 1 and 2).

The labeling system of IR rests on the premise that there are thresholds of difference above which two intervals are perceived to have categorical contrast. These thresholds account for the distinctions between symbols such as L and S or P^+ and $P^\sim$ and thus serve as the basis for categorical interval sizes here. In principle these thresholds vary depending on context; the values used in this study are given in Table 2.

TABLE 2
PROVISIONAL VALUES OF BOUNDARIES
BETWEEN IR'S ANALYTICAL CATEGORIES

Threshold	Value in music [1], [2]	Value in speech
L to S	6 semitones (st)	4 semitones (st)
R+ to R~	3 st	2.5 st
R~ to R0	N/A	.25 st
R~ to R-	3 st	2.5 st
P+ to P~	4 st	3 st
P~ to P0	N/A	.25 st
P~ to P-	4 st	3 st
Duration suppressing implication	1.5 times duration of previous tone	1.2 times duration of previous tone

In applying IR to speech, two additional difficulties present themselves. First, IR uses as input a series of discrete musical tones instead of speech's continuously varying pitch contour. Thus the continuous contour is converted to a stylized series of steady pitches, generally one per syllable, using weighted averages supplemented by human judgement. Second, the appropriate categorical thresholds appear to be larger for music than for speech. The values shown in Table 3 are

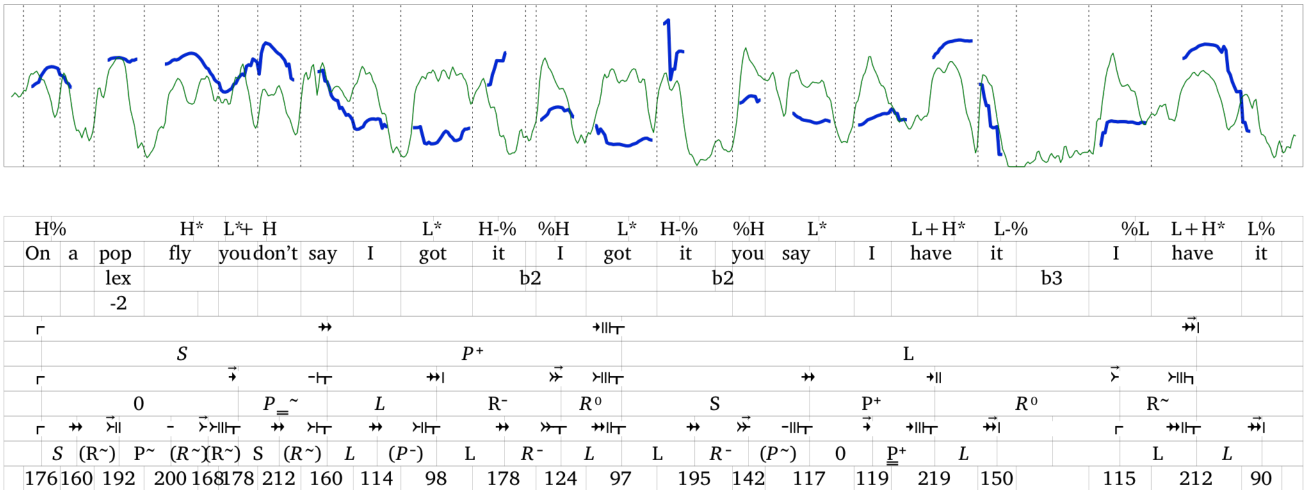


Fig. 3. College baseball coach quoted by broadcaster Vin Scully: "On a pop fly, you don't say, 'I got it, I got it,' you say, 'I have it, I have it.'" AM annotation appears above the words. IR analysis appears below the words. The numbers at bottom indicate syllables' stylized frequencies. Thick lines graph spoken pitch. Thin lines graph spoken intensity.

chosen because they generate plausible parsings. Analyses were generated with the help of scripts written by the author using Praat software [10].

III. Observations

Figs. 3–6, a selection of juxtaposed AM and IR analyses of spoken utterances, reveal a number of consistent points of comparison between AM and IR:

1. Tones that appear as suffix boundary tones (symbolized %) in AM appear as non-closed tones at shallow hierarchical levels in IR. (See figs. 3 and 6.)
2. In English and German, tones that AM treats as H* or L* tend to be closural on shallow hierarchical levels. Thus they emerge at deeper levels, where they can interact implicatively with tones from earlier and later speech segments. These implicative interactions correspond to the cohesive relationships between accented words as described by AM.
3. Where AM relies on pitch peaks and troughs to identify H* and L* accents, IR's employment of pitch and durational factors also results in robust identifications of these accents. This is even true in cases where the peak pitch occurs after the accented syllable, e.g. "YOU don't say" in Fig. 3 [11] and "malilita" in Fig. 4 [12]. While the AM analyses parse these two cases differently, IR has a single approach to both: the accented tones are initiations which emerge on deeper levels while the subsequent rise in pitch functions implicatively at the shallow level and thus does not figure at deeper levels.
4. While there are not always literal mappings between particular AM accent labels and particular IR labels, IR nevertheless generates robust descriptions of pitch and its functions that are consistent with AM's accounting of the features' functions. They are also more precise. AM's L+H* accent, whose sharply rising contour draws attention to contrasting information, is associated with

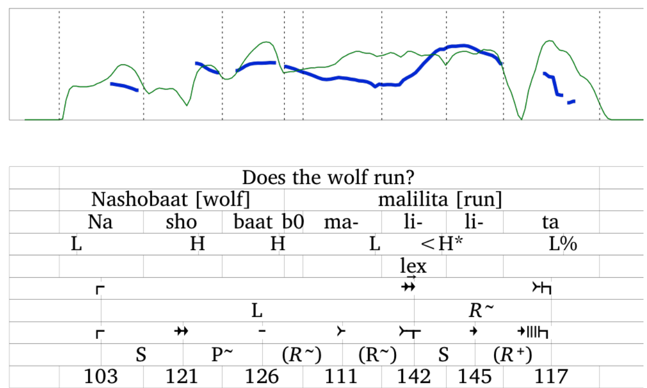


Figure 4. Chickasaw utterance: "Does the wolf run?"

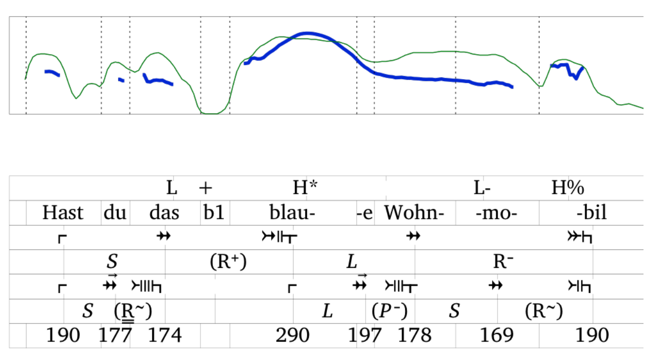


Figure 5. German utterance: "Is yours the blue camper?"

TABLE 3
AUTOSEGMENTAL-METRICAL SYMBOLS

Phonological Symbol	Meaning	Function Symbol	Meaning
H	high pitch	*	lexical accent
L	low pitch	–	phrase accent
^ or <	pitch upstep	%	boundary tone
!	pitch downstep		

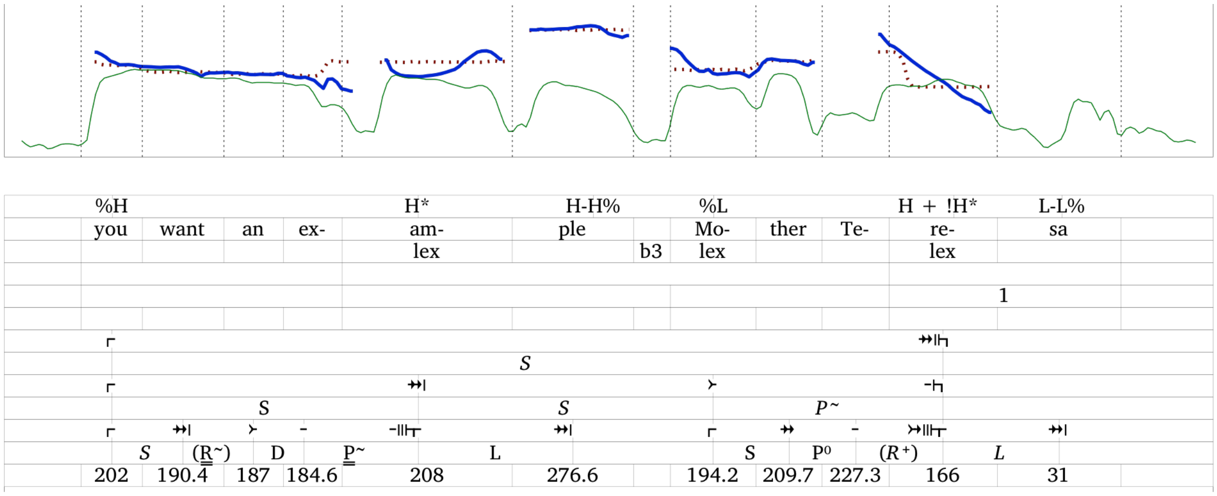


Figure 6. AM and IR analyses of the utterance “You want an example? Mother Teresa.” Dotted lines graph stylized pitch.

increasing interval size ⁺; hence the (R⁺) on “blaue” in Fig.5 [13] and the P⁺ on the first “I have it” in Fig. 3. The accent on “Teresa” in Fig. 6 [14], described as !H downstepped from the previous H in AM, is precisely specified in IR by being described as (*R*⁺), *P*~, and *S* in relation to earlier pitches on each of three levels.

5. IR captures resemblances between instances whose intonation differs superficially but coincides functionally. The German rising yes/no-question in Fig. 5 is seen to be similar in the IR analysis to the falling Chickasaw yes/no-question in Fig. 4. In these questions the answer is unknown to the speaker. IR can also make distinctions between instances whose intonation is superficially similar but differs functionally. In the rising yes/no question of Fig. 6, the speaker has already decided the answer, and IR offers an analysis different from those of Figs. 4 and 5.
6. IR is able to describe intonational mimesis of relationships between items in discourse. Symbols such as L and R⁺ denote intervals whose pitch evinces perceptual contrast; S and P⁻ denote non-contrast. Thus it is significant that the contrasting words “have” and “got” in Fig. 3 are linked by the contrastive interval L and that the non-contrastive relationship between the topics of the two halves of Fig. 6 is signaled by the non-contrastive *S*.

IV. Conclusion

In these analyses, IR is significant not merely for its assertions about expectation but also for its contributions to the study of segmentation, cohesion, and phonological categorization in sound streams. If IR successfully describes spoken intonation, it becomes possible to understand more deeply the musical analyses possible in IR, and thus the nature of music itself. IR’s interval-based approach may contribute usefully to the study of intonation. In addition, studies of speech-song hybrids such as chant, recitation, *Sprechmelodie*, parlando, and rap music would illuminate IR, and vice versa.

All of these aspects of IR draw attention to the expressive power of phonological processes that appear to be shared by music and language.

References

[1] E. Narmour, *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model*. Chicago: University of Chicago Press, 1990.

[2] E. Narmour, *The Analysis and Cognition of Melodic Complexity: The Implication-Realization Model*. Chicago: University of Chicago Press, 1992.

[3] E. Narmour, “Toward a Unified Theory of the I-R Model (Part I): Parametric Scales and their Analogically Isomorphic Structures,” *Music Perception*, vol. 33, no. 1, pp. 32–69, 2015.

[4] J. B. Pierrehumbert, “The phonology and phonetics of English intonation,” Ph.D. dissertation, Massachusetts Institute of Technology, 1980.

[5] J. Pierrehumbert and J. Hirschberg, “The Meaning of Intonational Contours in the Interpretation of Discourse,” in *Intentions in Communication*, P. R. Cohen, J. L. Morgan, and M. E. Pollack, Eds. MIT Press, 1990, pp. 271–312.

[6] A. Wennerstrom, *The Music of Everyday Speech: Prosody and Discourse Analysis*. Oxford: Oxford University Press, 2001.

[7] M. E. Beckman, J. Hirshberg, and S. Shattuck-Hufnagel, “The Original ToBI System and the Evolution of the ToBI Framework,” in *Prosodic Typology: The Phonology of Intonation and Phrasing*, S.-A. Jun, Ed. Oxford: Oxford University Press, 2005, pp. 9–54.

[8] D. R. Ladd, *Intonational Phonology*, 2nd ed. Cambridge: Cambridge University Press, 2008.

[9] E. H. Margulis, “A Model of Melodic Expectation,” *Music Perception*, vol. 22, no. 4, pp. 663–714, Summer 2005.

[10] P. Boersma and D. Weenink, *Praat: Doing Phonetics by Computer [computer program]*. 2014.

[11] “Vin Scully college grammar story,” *YouTube*, Sep. 13, 2011 [Video file]. Available: https://www.youtube.com/watch?v=1doxB2AZnaI. [Accessed: April 23, 2016].

[12] M. Gordon, “An autosegmental-metrical model of Chickasaw intonation,” in *Prosodic Typology: The Phonology of Intonation and Phrasing*, S.-A. Jun, Ed. Oxford: Oxford University Press, 2005, pp. 301–330.

[13] M. Grice, S. Baumann, and R. Benzmüller, “German Intonation in Autosegmental-Metrical Phonology,” in *Prosodic Typology: The Phonology and Intonation and Phrasing*, S.-A. Jun, Ed. Oxford: Oxford University Press, 2005, pp. 55–83.

[14] Alejna Brugos, Stefanie Shattuck-Hufnagel, and Nanette Veilleux, “Transcribing Prosodic Structure of Spoken Utterances with ToBI.” MIT Open Courseware, ch. 2.7. [Accessed: Nov. 15, 2015].

Acoustic and musical foundations of the speech/song illusion

Adam Tierney,*[1] Aniruddh Patel#[2], Mara Breen^[3]

*Department of Psychological Sciences, Birkbeck, University of London, United Kingdom

#Department of Psychology, Tufts University, United States

^Department of Psychology and Education, Mount Holyoke College, United States

[1]a.tierney@bbk.ac.uk, [2]a.patel@tufts.edu, [3]mbreen@mtholyoke.edu,

ABSTRACT

In the 'speech-to-song illusion', certain spoken phrases sound like song when isolated from context and played repeatedly. Previous work has shown that this perceptual transformation occurs more readily for some phrases than others, suggesting that the switch from speech to song perception depends in part on certain cues. We conducted three experiments to explore how stimulus characteristics affect the illusion. In Experiment 1, we presented 32 participants with a corpus of 24 spoken phrases which become more song-like when repeated and 24 spoken phrases which continue to sound like speech when repeated. After each of 8 repetitions participants rated the extent to which the phrase sounded like song versus speech. Regression modeling indicated that an increase in song perception between the first and eighth repetitions was predicted by a) greater stability of the pitches within syllables, b) a better fit of average syllable pitches to a Bayesian model of melodic structure, and c) less variability in beat timing, as extracted by a beat-tracking algorithm. To investigate whether pitch characteristics play a causal role in the speech-to-song transformation, we elicited ratings of the stimuli from Experiment 1 after manipulating them to have larger pitch movements within syllables (Experiment 2, n = 27) or to have average pitches of syllables which resulted in poorer melodic structure (Experiment 3, n = 31). Larger pitch movements within syllables did not decrease the size of the illusion compared to Experiment 1; however, the illusion was significantly weaker when the intervals between pitches were altered. These results suggest that the strength of the illusion is determined more by pitch relationships between than within syllables, such that phrases with pitch relationships between spoken syllables that resemble those of Western tonal music are more likely to perceptually transform than those that do not.

I. INTRODUCTION

Music can take on a wide variety of forms, even within Western culture. Musical genres are marked by large differences in the timbral, rhythmic, melodic, and harmonic patterns upon which composers and musicians draw. Given this variety of musical subcultures, one might expect there to be little agreement across the general population as to the qualities that cause a sound sequence to be perceived as more or less musical.

However, there exist certain spoken recordings that listeners tend to perceive as sounding like song when isolated from context and repeated (Deutsch 2011). Across subjects, this transformation is stronger for some recordings than others (Tierney et al. 2013), and musicians and non-musicians agree as to which examples do and do not transform (Vanden Bosch der Nederlanden et al. 2015a, 2015b). The existence of this illusion suggests that there are certain cues on which listeners tend to rely when judging the musicality of a linguistic phrase, and that these cues are relatively unaffected by musical experience. The fact that repetition is necessary for the speech-song transformation to take place also suggests that a certain amount of time is necessary for the detection of at least a subset of these cues.

What, then, are the minimal characteristics which need to be present in order for a linguistic phrase to sound musical? One explanation—the acoustic cue theory--is that these characteristics are primarily acoustic in nature. For example, the speech-to-song transformation takes place more often for stimuli that have relatively flat pitches within syllables (Tierney et al. 2013, Falk et al. 2014). This may facilitate the categorization of pitches into pitch classes, enabling listeners to perceive the pitch sequence as a melody—with the result that the pitches that listeners perceive are distorted from the original pitches of the sequence (Vanden Bosch der Nederlanden et al., 2015b). Thus, listeners may tend to hear any sequence of flat pitches as musical. Supporting this account is the fact that random sequences of pitches are rated as more musical if they have been repeated (Margulis and Simchy-Gross 2016). Furthermore, repetition may be necessary for the illusion to take place because exact repetition causes speech perception resources to become satiated, allowing musical perception to take over. This would explain why the speech-song transformation is stronger for languages that are more difficult for a listener to pronounce (Margulis et al. 2015).

A second possibility—the musical cue theory—is that basic acoustic cues such as flat pitches within syllables are necessary but not sufficient for the perception of a spoken sentence as song. According to this account the speech-to-song illusion would only occur for spoken sentences which feature both acoustic pre-requisites such as pitch flatness and musical cues matching certain basic characteristics of Western music. For example, a sequence featuring relatively flat pitches but an abundance of tritone intervals may be unlikely to be perceived as music, because tritone intervals are rare in Western music. This theory offers an alternate (but not exclusive) explanation for the necessity of repetition for eliciting the illusion, as a certain amount of time may be necessary for the detection of some or all of these musical cues. This theory suggests not only that listeners across the general population can make relatively sophisticated musical judgments (Bigand and Poulin-Charronnat 2006), but that they can make these judgments about non-musical stimuli.

Here we tested the musical cue theory of the speech-song illusion using the speech-song corpus first reported by Tierney et al. (2013). This corpus consists of 24 stimuli which listeners perceive as song after repetition (Illusion stimuli) and 24 stimuli which listeners perceive as speech after repetition (Control stimuli). We repeated each stimulus eight times and asked a new set of listeners with a relatively small amount of musical training to rate how much the stimulus sounded like song after each repetition. We measured the musical characteristics of each stimulus using computational models of melodic structure (Temperley 2007) and musical beat structure (Ellis 2007). Our prediction was that these musical characteristics would explain additional variance in the extent to which each stimulus transformed into song with repetition, even after syllable pitch flatness was accounted for. Furthermore, we predicted that musical characteristics would correlate with the change in song ratings due to repetition but not with song ratings after a single repetition.

II. Experiment 1

A. Methods

1) Participants. 32 participants were tested (14 male). The mean age was 33.7 (standard deviation 9.4). The mean amount of musical training was 1.9 years (standard deviation 3.8).

2) Stimuli. Stimuli consisted of the 24 Illusion stimuli and 24 Control stimuli described in Tierney et al. (2013). These were short phrases (mean 6.6 (sd 1.5) syllables) extracted from audiobooks.

3) Procedures. Participants were recruited via Mechanical Turk, a website which enables the recruitment of participants for internet-based work of various kinds. This study was conducted online using the Ibex paradigm for internet-based psycholinguistics experiments (http://spellout.net/ibexfarm). Participants were told that they would hear a series of spoken phrases, each repeated eight times. After each presentation they were given three seconds to indicate, on a scale from 1 to 10, how much the stimuli sounded like song versus like speech. Judgments could be made either by clicking on boxes which contained the numbers or by pressing the number keys. Order of presentation of the stimuli was randomized. Participants also heard, interspersed throughout the test, four "catch" trials in which the stimulus actually changed from a spoken phrase to a sung phrase between the fourth and fifth repetitions. Data from participants whose judgments of the sung portions of the catch trials were not at least 1.5 points higher than their judgments of the spoken portions were excluded from analysis. This constraint resulted in the exclusion of 8 participants from Experiment 1, 7 participants from Experiment 2, and 8 participants from Experiment 3. Excluded participants are not included in the subject counts of demographic descriptions within the Participants section for each Experiment.

4) Analysis. To minimize the influence of inter-subject differences on baseline ratings, ratings were normalized prior to analysis with the following procedure: each subject's mean rating across all repetitions for all stimuli was subtracted from each data point for that subject. Data were then averaged across stimuli within a single subject. This generated, for each subject, average scores for each repetition for Illusion and Control stimuli. A repeated measures ANOVA with two within-subjects factors (condition, two levels; repetition, eight levels) was then run; an interaction between repetition and condition was predicted, indicating that the Illusion stimuli transformed to a greater extent than the Control stimuli after repetition.

To investigate the stimulus factors contributing to the speech-song illusion, for each stimulus initial and final ratings were averaged across all subjects. The initial rating and the rating change between the first and last repetition were then calculated and correlated with several stimulus characteristics. First, we measured the average pitch flatness within syllables by manually marking the onset and offset of each syllable, tracking the pitch contour using autocorrelation in Praat, calculating the change in pitch between each time point (in fractions of a semitone), and then dividing by the length of the syllable. Thus, pitch flatness was measured in semitones per second.

Second, we determined the best fit of the pitches in each sequence to a diatonic key using a Bayesian key-fitting algorithm (Temperley, 2007) which evaluates the extent to which pitch sequences fit characteristics of melodies from western tonal music by assessing them along a number of dimensions, including the extent to which the distribution of interval sizes fits the interval size distribution of Western vocal music and fit to tonal key profiles.

The model considers four sets of probabilities—*key profile, central pitch profile, range profile,* and *proximity profile*—which are all empirically generated from the Essen Folksong Collection, a corpus of 6217 European folk songs. The key profile is a vector of 12 values indicating the probability of occurrence for each of the 12 scale tones in a melody from a specific key, normalized to sum to 1. For example, on average 18.4% of the notes in a melody in a Major key are scale degree 1 (e.g., C in C Major), while only 0.1% are scale degree #1 (e.g., C# in C Major). The central pitch profile (c) is a normal distribution of pitches represented as integers (C4 = 60) with a mean of 68 (Ab4) and variance of 13.2, which captures the fact that melodies in the Essen corpus are centered within a specific pitch range. The range profile is a normal distribution with a mean of the first note of the melody, and variance of 29, which captures the fact that melodies in the Essen corpus are constrained in their range. The proximity profile is a normal distribution with a mean of the previous note, and variance of 8.69, which captures the fact that melodies in the Essen corpus have relatively small intervals between adjacent notes. The final parameter of the model is the *RPK profile*, which is calculated at each new note as the product of the key profile, range profile, and proximity profile. The inclusion of the RPK profile captures the fact that specific notes are more probable after some tones than others.

Calculating the probability of each of the 24 (major and minor) diatonic keys given a set of notes is done using the equation below:

$$P(pitch\ sequence) = \sum_{k,c}\left(P(k)P(c)\prod_{n} RPK_n\right)$$

P(k) is the probability of any key (k) being chosen (higher for major than minor keys), *P(c)* is the probability of a central pitch being chosen, and RPK_n is the RPK profile value for all pitches of the melody given the key, central pitch, and prior pitch for each note. We defined melodic structure as the best fit of each sequence to the key (k) that maximized key fit in the equation.

Finally, we used a computer algorithm designed to find beat times in music (Ellis 2007) to determine the location of each beat, and then we calculated the standard deviation of inter-beat times to measure beat regularity.

The beat tracking algorithm works as follows. First, it divides a sound sequence into 40 equally-spaced Mel frequency bands, and extracts the onset envelope for each band by taking the first derivative across time. These 40 onset envelopes are then averaged, giving a one-dimensional vector of onset strength across time. Next, the global tempo of the sequence is estimated by taking the autocorrelation of the onset envelope, then multiplying the result by a Gaussian weighting function. Since we did not have a strong a priori hypothesis for the tempo of each sequence, the center of the Gaussian weighting function was set at 120 BPM, and the standard deviation was set at 1.5 octaves. The peak of the weighted autocorrelation function is then chosen as the global tempo of the sequence. Finally, beat onset times are chosen using a dynamic programming algorithm which maximizes both the onset strength at chosen beat onset times and the fit between intervals between beats and the global tempo. A variable called “tightness” sets the relative weighting of onset strength and fit to the global tempo; this value was set to 100, which allowed a moderate degree of variation from the target tempo.

The beat times chosen by this algorithm tend to correspond to onsets of stressed syllables, but can also appear at other times (including silences) provided that there is enough evidence for beat continuation from surrounding stressed syllable onsets. This algorithm permits non-isochronous beats, and is therefore ideal for extracting beat times from speech, despite the absence of metronomic regularity (Schultz et al. 2015). As this procedure was only possible when phrases contained at least three identifiable beats, we were only able to calculate beat variability for 42 of the 48 stimuli. As a result, correlational and regression analyses were only run on these 42 stimuli.

B. Results

Song ratings increased with repetition across all stimuli (main effect of repetition, $F(1.5, 48.0) = 47.9$, $p < 0.001$). However, this increase was larger for the Illusion stimuli (interaction between condition and repetition, $F(1.5, 46.8) = 62.7$, $p < 0.001$). Moreover, song ratings were greater overall for the Illusion stimuli compared to the Control stimuli (main effect of condition, $F(1, 31) = 83.6$, $p < 0.001$). (See Figure 1, black lines, for a visual display of song ratings across repetitions for Illusion and Control stimuli in Experiment 1.)

Table 1 displays correlations between initial ratings and rating changes, and the three stimulus attributes. After a single repetition, subjects’ reported song perception was only correlated with beat variability. However, rating change was correlated with pitch flatness, melodic structure, and beat variability.

Table 1. Correlation between song ratings and stimulus characteristics. Bolded cells indicate significance at p < 0.05 (Bonferroni corrected).

r-values	Initial rating	Rating change
Pitch flatness	0.151	**0.588**
Melodic structure	0.141	**0.541**
Beat variability	**0.456**	**0.402**

We used hierarchical linear regression to determine whether pitch flatness, melodic structure, and beat variability contributed independent variance to song ratings. By itself, pitch flatness predicted 34.6% of variance in song ratings ($\Delta R^2 = 0.346$, $F = 21.2$, $p < 0.001$). Adding melodic structure increased the variance predicted to 45.2% ($\Delta R^2 = 0.105$, $F = 7.5$, $p < 0.01$). Adding beat variability increased the variance predicted to 55.6% ($\Delta R^2 = 0.105$, $F = 7.5$, $p < 0.01$).

C. Discussion

As reported previously (Tierney et al. 2013), stimuli for which the speech-song transformation was stronger had flatter pitch contours within syllables. However, this characteristic was not sufficient to fully explain why some stimuli transformed more than others. Adding musical cues, namely melodic structure and beat variability, improved the model. This result suggests that listeners who are relatively musically inexperienced rely on melodic and rhythmic structure when judging the musical qualities of spoken phrases.

We also found, contrary to our prediction, that musical beat variability correlated not just with the increase in song perception with repetition but also with song ratings after a single repetition. Melodic structure and pitch flatness, however, only correlated with song ratings after repetition. This result suggests that the rhythmic aspects important for the perception of song can be perceived immediately, while the melodic aspects take time to extract. This finding could provide a partial explanation for why repetition is necessary for the speech-song illusion to take place, but does not rule out the possibility that satiation of speech perception resources is playing an additional role.

Although these results suggest that syllable pitch flatness, melodic structure, and beat variability all influence speech-song transformation, our correlational design was unable to show that these factors played a causal role. To begin to investigate this issue we ran two follow-up experiments in which syllable pitch flatness (Experiment 2) and melodic structure (Experiment 3) were experimentally varied. We predicted that the removal of either of these cues would diminish the speech-song effect.

III. Experiment 2

D. Methods

1) Participants. 27 participants were tested (17 male). The mean age was 33.1 years (standard deviation 7.4). The mean amount of musical training was 1.0 years (standard deviation 1.6).

2) Stimuli. Illusion and Control stimuli from Experiment 1 were altered for Experiment 2. First, the ratio of pitch variability within syllables in the Control stimuli and Illusion stimuli was measured. On average pitch variability within syllables was 1.49 times greater for Control stimuli than for Illusion stimuli. This ratio was then used to alter the pitch contours of syllables (using Praat) such that pitch movements within syllables for Illusion stimuli were multiplied by 1.49, while pitch movements within syllables for Control stimuli were divided by 1.49. This process switched the pitch flatness of Control and Illusion stimuli while maintaining other characteristics such as beat variability and melodic structure.

3) Procedures. Same as Experiment 1.

4) Analysis. To determine whether the pitch flatness manipulation altered song perception ratings, data from Experiment 1 and Experiment 2 were compared using a repeated measures ANOVA with two within-subject factors (condition, two levels; repetition, eight levels) and one between-subject factor (experiment).

E. Results

Similar to the data from Experiment 1, across all stimuli for both experiments there was an increase in song ratings with repetition (main effect of repetition, $F(1.4, 81.9) = 82.2$, $p < 0.001$), although this increase was larger for the Illusion stimuli (interaction between condition and repetition, $F(1.6, 88.6) = 99.9$, $p < 0.001$). Song ratings were also greater overall for the Illusion stimuli compared to the Control stimuli (main effect of condition, $F(1,57) = 133.9$, $p < 0.001$). However, the pitch flatness manipulation had no measurable effect on song perception ratings (no interaction between experiment and repetition, $F(1,57) = 1.0$, $p > 0.1$; no interaction between experiment, repetition, and condition, $F(1.6, 88.6) = 0.56$, $p > 0.1$). See Figure 1 for a visual comparison between song ratings from Experiment 1 and from Experiment 2.

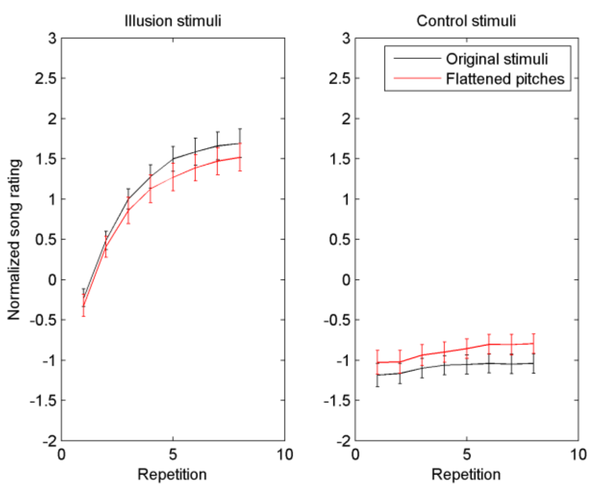


Figure 1. Song ratings for Illusion and Control stimuli, unaltered (black line) and with increased pitch variation for the Illusion stimuli and decreased pitch variation for the Control stimuli (red line).

F. Discussion

Contrary to our predictions, switching the syllable pitch flatness characteristics of the Illusion and Control stimuli did not affect the magnitude of the speech-song transformation. Falk et al. (2014), on the other hand, found that increasing tonal target stability boosted the speech-song transformation; however, our pitch manipulations in this study were much smaller than those used by Falk et al. (2014). While these results do not rule out a role for syllable pitch flatness entirely, they do suggest that this factor does not play a major role in distinguishing transforming from non-transforming stimuli in this particular corpus.

IV. Experiment 3

G. Methods

1) Participants. 31 participants were tested (21 male). The mean age was 35.5 years (standard deviation 7.5). The mean amount of musical training was 1.1 years (standard deviation 2.3).

2) Stimuli. Illusion and Control stimuli from Experiment 1 were altered for Experiment 3 using a Monte Carlo approach with 250 iterations across each of the 48 stimuli. For each iteration the pitch of each syllable was randomly shifted to between 3 semitones below and 3 semitones above its original value. Temperley's (2007) algorithm was then used to determine which of the 250 randomizations resulted in the worst fit to Western melodic structure, and this randomization was used to construct the final stimulus. Note that this manipulation does not affect the pitch flatness within syllables.

3) Procedures. Same as Experiment 1.

4) Analysis. To determine whether the melodic structure manipulation altered song perception ratings, data from Experiment 1 and Experiment 3 were compared using a repeated measures ANOVA with two within-subject factors (condition, two levels; repetition, eight levels) and one between-subject factor (experiment).

H. Results

Similar to the data from Experiment 1, across all stimuli for both experiments there was an increase in song ratings with repetition (main effect of repetition, $F(1.4, 86.0) = 81.6$, $p < 0.001$), although this increase was larger for the Illusion stimuli (interaction between condition and repetition, $F(1.5, 91.2) = 105.6$, $p < 0.001$). There was also a tendency for the Illusion stimuli to sound more song-like overall (main effect of condition, $F(1, 62) = 161.9$, $p < 0.001$). Importantly, however, the melodic structure manipulation changed song perception ratings: the repetition effect for Illusion stimuli was larger for the original stimuli than for the altered stimuli (3-way interaction between repetition, condition, and experiment, $F(1.5, 91.2) = 6.1$, $p < 0.01$). The melodic structure manipulation also had different overall effects for the two classes of stimuli, decreasing song perception ratings for the Illusion stimuli but increasing song perception ratings for the Control stimuli (interaction between condition and experiment, ($F(1, 62) = 6.3$, $p < 0.05$). See Figure 2 for a visual comparison between song ratings from Experiment 1 and from Experiment 3.

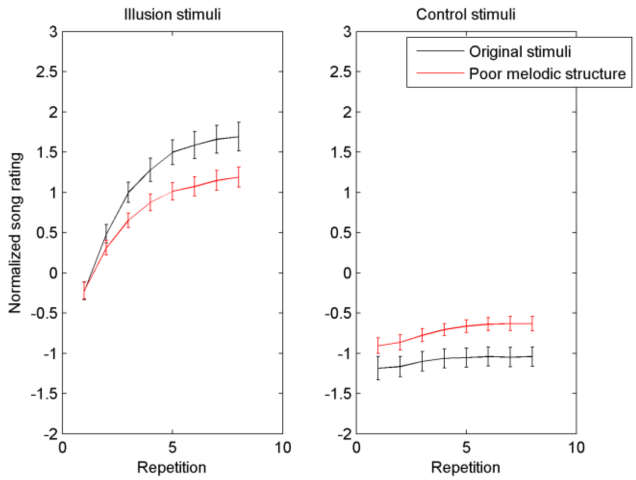


Figure 2. Song ratings for Illusion and Control stimuli, unaltered (black line) and altered to poorly fit a Bayesian model of melodic structure (red line).

I. Discussion

As predicted, forcing the stimuli to more poorly fit a model of melodic structure diminished the magnitude of the speech-song transformation. These results suggest that melodic structure plays a causal role in the speech-song transformation. However, initial song perception ratings were unaffected by the melodic structure manipulation. This, along with the results of Experiment 1, further supports the idea that the increase in song perception with repetition is due in part to a gradual extraction of melodic information from the sequence. Unexpectedly, the melodic fit manipulation also increased song perception for the Control stimuli. We do not currently have a theoretical framework for this finding and so further investigation is needed to pinpoint the source of this effect.

V. General Discussion

We found that within-syllable pitch contour flatness, melodic structure, and beat variability predicted the magnitude of the speech-song transformation. Of these characteristics, only beat variability predicted song ratings after a single repetition. This result suggests that the melodic aspects of a sound sequence take time to extract and could explain why the speech-song illusion requires repetition, as well as why the repetition of random tone sequences increases judgments of their musicality (Margulis and Simchy-Gross 2016). The rhythmic aspects of a sound sequence, on the other hand, may be immediately accessible. If so, one prediction is that a non-tonal rhythmic sequence may not increase in musicality with repetition.

Altering within-syllable pitch flatness did not modulate the speech-song illusion. However, decreasing the extent to which a sequence fit a Bayesian model of melodic structure decreased the intensity of the speech-song illusion. It is possible that the correlation between syllable pitch flatness and speech-song transformation is driven by a third variable. For example, recordings with more overall pitch movement may have both less flat pitch contours within syllables and larger pitch intervals across syllables, and this second factor may be a more important cue for speech-song transformation. Future work in which the rhythmic properties of the stimuli are altered could determine whether beat variability plays a causal role in the speech-song illusion. We predict that increasing beat variability will diminish both initial song perception ratings and the increase in song perception with repetition.

As a whole, our results suggest that musical characteristics of stimuli such as melodic structure and beat variability may be more important than acoustic characteristics such as pitch variability within syllables in determining the strength of the speech-song illusion. If true, this may enable the creation of stimuli that are closely matched on acoustic characteristics, differing only in those musical characteristics necessary for eliciting the speech-song illusion. Such stimuli would be ideal for comparing the neural correlates and perceptual consequences of speech and music perception. Furthermore, our results add to the growing body of work demonstrating that musical sophistication is widespread in the general population (Bigand and Poulin-Charronnat 2006). Indeed, these findings suggest that listeners not only possess sophisticated musical knowledge, they can apply this knowledge to judge the musicality of sound sequences that were never intended to be heard as music.

REFERENCES

Bigand, D., & Poulin-Charronnat, B. (2006). Are we "experienced listeners"? A review of the musical capacities that do not depend on formal training. *Cognition, 100,* 100-130.

Deutsch, D., Henthorn, T., & Lapidis, R. (2011). Illusory transformation from speech to song. *Journal of the Acoustical Society of America, 129,* 2245-2252.

Ellis, D. (2007). Beat tracking by dynamic programming. *Journal of New Music Research, 36,* 51-60.

Falk, S., Rathcke, T., & Dalla Bella, S. (2014). When speech sounds like music. *Journal of Experimental Psychology: Human Perception and Performance, 40,* 1491-1506.

Margulis, E., Simchy-Gross, R., & Black, J. (2015). Pronunciation difficulty, temporal regularity, and the speech-to-song illusion. *Frontiers in Psychology, 6,* 48.

Margulis, E., & Simchy-Gross, R. (2016). Repetition enhances the musicality of randomly generated tone sequences. *Music Perception, 33,* 509-514.

Schultz, B., O'Brien, I., Phillips, N., McFarland, D., Titone, D., & Palmer, C. (2015). *Applied Psycholinguistics.* DOI: 10.1017/S0142716415000545.

Temperley, D. (2007). *Music and probability*. Cambridge, MA: MIT Press.

Vanden Bosch der Nederlanden, C., Hannon, E., & Snyder, J. (2015a). Everyday musical experience is sufficient to perceive the speech-to-song illusion. *Journal of Experimental Psychology: General, 144,* e43-e49.

Vanden Bosch der Nederlanden, C., Hannon, E., & Snyder, J. (2015b). Finding the music of speech: musical knowledge influences pitch processing in speech. *Cognition, 143,* 135-140.

Finding the beat in poetry: The role of meter, rhyme, and lexical content in speech production and perception

Benjamin G. Schultz[1] & Sonja A. Kotz[1]
[1]Basic & Applied Neuro Dynamics Laboratory, Maastricht University, Department of Neuropsychology & Psychopharmacology, Faculty of Psychology & Neuroscience, Maastricht University, The Netherlands

Poetry is an artistic form of communication that can contain meter, rhyme, or lexical content. We examined how the temporal regularity of spoken German poetry varied as a function of these features. Using a beat tracker that was previously used to measure speech rate based on moments of acoustic stress, we examined how stress rate and variability were influenced by meter, rhyme, and lexical content during read poetry performances. Theories of cognitive fluency state that less cognitively demanding tasks are performed more quickly and less variably than more cognitively demanding tasks. Based on evidence that meter, rhyme, and lexical words would increase cognitive fluency, we hypothesized that the stress rate would be faster and less variable for: 1) poems with an implied meter compared to those without, 2) rhyming poems compared to non-rhyming poems, and 3) poems with lexical words compared to pseudo-words. A German poet performed poetry readings consisting of single phrases where the written stanzas varied in meter (meter, no meter), rhyme (rhyme, no rhyme), and lexical content (lexical words, pseudo words). Recordings were subjected to beat tracking analyses that yielded the mean and standard deviation of the inter-beat intervals for each stanza measuring the stress rate and variability, respectively. In line with our hypothesis, poems with a meter were spoken at a faster rate and with less variability than poems without meter (*ps* < .001). Bayes factor t-tests revealed substantial evidence for the null hypothesis indicating no difference in stress rate or stress variability for the factors rhyme and lexical content. Results indicate that, during readings of poetry, meter increases stress rate and decreases variability, suggesting that an implied meter may increase cognitive fluency. However, rhyme and lexical content do not influence stress rate or variability. The perception of the poetry readings will be examined using recollection tasks and EEG recordings.

Vowel perception by congenital amusics

Jasmin Pfeifer[1,2], Silke Hamann[1]
[1]Amsterdam Center for Language and Communication, University of Amsterdam, Amsterdam, The Netherlands
[2]Institute for Language and Information, Heinrich-Heine-University, Düsseldorf, Germany

Congenital amusia is a neurodevelopmental perception disorder that has mostly been investigated in regard to difficulties in music perception (Peretz et al. 2002). Recent studies, however, show that amusics also have an impaired perception of language, e.g. difficulties in differentiating intonation (i.e. the rise and fall of pitch over a phrase; Patel et al. 2008) or lexical pitch contrasts in tone languages (such as Mandarin; Liu et al. 2015) .

The present study tests German amusics' perception of vowels, focusing on durational and qualitative (i.e. spectral) characteristics. We hypothesize that vowel pairs with small differences in duration or quality are difficult to discriminate by amusics, and that amusics employ a different cue weighting of spectral and durational cues compared to normal listeners. Furthermore, we hypothesize that amusics have severe problems perceiving vowel differences if these are presented with a short inter-stimulus interval (ISI), as amusics have shown problems with rapid auditory temporal processing (Williamson et al. 2010).

We tested this with vowel stimuli synthesized on the basis of the German front mid vowels /e:/ (long and tense) and /ɛ/ (short and lax): By varying the spectral (tense vs. lax) as well as durational information (long vs. short), we created four separate continua (see e.g. Boersma & Escudero 2005). These continua were used in a counterbalanced AXB task, where the ISI was once 0.2 s and once 1.2 s.

Preliminary findings show that the amusics' overall performance is worse than that of the controls, and that they have severe problems with an ISI of 0.2 s (almost chance performance).

This study thus supports earlier findings that amusia is not music-specific but affects general auditory mechanisms that are also relevant for language. In addition, it shows that these impairments affect the perception of vowel quality and duration.

A tool for the quantitative anthropology of music: Use of the nPVI equation to analyze rhythmic variability within long-term historical patterns in music

Joseph R Daniele
University of California, Berkeley

ABSTRACT: The development of musical style across time and geography is of particular interest to historians and musicologists yet quantitative evidence to support these trends has been lacking. This paper illustrates a novel application of the nPVI ('normalized pairwise variability index') equation to probe and quantify the rhythmic components of music over time. The nPVI equation quantifies the average difference between adjacent events in a sequence (e.g. musical notes in a melody, successive vowels in a spoken sentence). Building upon an earlier finding that German/Austrian composer nPVI values increased steadily from 1600 to 1950 (while Italian composers showed no such increase), the nPVI 'distribution' of themes from individual composers was quantitatively explored. Interestingly, the proportion of 'low nPVI' or 'Italianate' themes decreases rapidly with time while 'high nPVI' (more Germanic) themes concomitantly increase in frequency. 'Middle range nPVIs' exhibit a constant incidence, arguing for a replacement of 'low nPVIs' (Italianate) with 'high nPVIs' over a short time instead of a more modest, long-term progressive shift. Thus, the precise rhythmic components of complex stylistic shifts in music can be quantitatively extracted from music and support the historical record and theory.

1. BACKGROUND

The nPVI ('normalized pairwise variability index') is an equation originally developed for linguistic analysis that quantifies the average amount of durational contrast between successive events in a sequence (e.g. musical notes in a melody or successive vowels in a spoken sentence). A high nPVI value, for instance, indicates greater durational contrast between adjacent events in a sequence (cf. the Appendix from [1] for an example of nPVI computation). Originally developed by phoneticians to show that "stress-timed" languages (British English, German, Dutch) had naturally higher nPVI values than "syllable-timed" languages (French, Spanish, Italian), Patel and Daniele [2] used the nPVI to show that a composer's native language directly influences the rhythms he/she writes[3]–[5] (for recent data on vocalic nPVI in English, German, Italian, and Spanish, see Arvaniti 2012, Figure 2b [6]). Recent studies of the nPVI have begun to explore the variation of this statistic with respect to time and culture. More specifically, Daniele and Patel have demonstrated that the nPVI can reveal patterns in the historical analysis of music [1], [7]–[9]. By plotting mean nPVI vs. midpoint year for German/Austrian and Italian composers who lived between ~1600 and 1950 a steady increase in nPVI values was observed over this period for German/Austrian music while Italian music showed no such increase [1]. These data are consistent with the idea (from historical musicology) that the influence of Italian music on German music began to wane in the second half of the 1700s due to a rise of musical nationalism in Germany [10]. Of note, these findings were replicated and expanded upon to include an initial increase and decrease in music from 34 French composers during this era [11].

It is important to note that the analyses performed by Daniele & Patel, e.g. [1], [8], were based on assigning each composer a single, average nPVI value. More recently, when composer's lives were demarcated into different compositional epochs (by historical musicologists) [7] it was found that the mean nPVI (for each compositional period) does not vary dramatically. Nonetheless, many questions remain. For instance, what is the rhythmic "distribution" of each composer during this transition from 1760-1800? To what extent might historical factors (the "Italian" influence) affect nPVI distribution over time?

2. AIMS

To test if nPVI could be used to detect dramatic historical shifts in music, the change in nPVI distribution was explored with respect to the aforementioned 1760-1800 stylistic transition in German music and compared to a culture where no change in nPVI would be expected (e.g. Italian). A simulation in which nPVI was modeled to increase modestly over a long period of time was then compared to the calculated data. Thus, this paper illustrates the utility of the nPVI equation and method in quantifying the precise rhythmic underpinnings of a complex stylistic transition in music history.

3. METHOD

Theme List, exclusion Criteria, and Data Analysis**:** Data collection and analysis was performed according to the previous methods and exclusion criteria were enumerated extensively in **[7], [1]**. As in the Daniele & Patel studies, the musical materials for the current work were drawn from *A Dictionary of Musical Themes, Revised Edition* **[12]**. The midpoint year (the mathematical average of the birth and death year, representing when the composer was active) was considered in grouping these individuals into periods from the "Baroque/Classical Era" (1600-1750/1750-1810) which took place before and during the 1760-1800 transition to more stylistically "German" music and the "Romantic Era" (1825-1900) which followed this transition (**[13]**, Vol. 9 pp. 708-744). Beethoven was considered one of the Romantic composers despite writing before the "Romantic period" **[14], [15]**. Modeling of nPVI distributions over time was completed using the "Normal Distribution Probability Calculator" in Sigma XL. More explicitly, distributions were calculated by plugging in composer mean nPVI, standard deviation, and nPVI range into the aforementioned program and then plotting the resulting proportions for each range.

4. RESULTS

With the intent to chronicle the rhythmic underpinnings of German musical evolution the median nPVI of German/Austrian composers from different eras of musical history ("Baroque/Classical" 1600-1825, and "Romantic" ~1825-1900) ([13]; and see Method) was calculated and used to categorize composers. Since these grouped data sets ("Baroque/Classical" composers (Bach, Haydn, Mozart)) and "Romantic" composers (Beethoven through R. Strauss, see **Table 1**) were deemed "non-normal" using the Anderson Darling Test for Normality, the medians (and not means) were used (nPVI = 32.2 for the "Baroque/Classical" period and nPVI = 43.0 for "Romantic"). These values were used to demarcate bins for each composer's theme corpus (**Table 1**). It is of note to mention that the nPVI at the intercept of the German/Austrian and Italian slopes [1] was 41.3 (for a 15 theme cutoff for Italians). Thus, the use of an nPVI of 43.0 as a demarcation point from the "Italian" style to the "German" style is reasonable.

Table 1

Composer	# themes	Mean nPVI	Midpoint Year	% themes w/ nPVI < 32.2	% themes w/ 32.2>nPVI>43	% themes w/ nPVI>43	Natio-nality
J.S. Bach	351	27.5	1717.5	65.8	16.2	17.9	GER
Haydn	278	35.8	1770.5	48.9	21.2	29.9	AUT
Mozart	460	41.7	1773.5	39.1	17.8	43.0	AUT
Beethoven	487	43.0	1798.5	40.5	15.4	44.1	GER
Schubert	232	47.1	1812.5	28.0	22.8	49.1	AUT
Mendelssohn	145	39.9	1828	43.4	13.1	43.4	GER
Schumann	216	41.2	1833	42.1	13.9	44.0	GER
Wagner	82	63.9	1848	11.0	11.0	78.0	GER
J. Strauss, Jr.	95	55.1	1862	18.9	14.7	66.3	AUT
Brahms	362	43.5	1865	34.5	18.8	46.7	GER
R. Strauss	111	60.0	1906.5	15.3	10.8	73.9	GER

When the frequencies in **Table 1** were plotted against each composer's midpoint year, and regression lines fitted, a striking trend was observed (**Figure 1**). The proportion of 'low nPVIs' (<32.2) significantly decreased with time (Frequency (%) = -0.25*(Midpoint Year) +492.26, $R^2 = 0.68$, $p < 0.01$) while the proportion of themes with higher nPVIs (>43) significantly increased with time (Frequency (%) = 0.29*(Midpoint Year) - 471.74, $R^2 = 0.71$, $p < 0.01$). No significant change in the proportion of themes with nPVI 32.1 to 43.0 was seen (p=0.14) suggesting the majority of composers wrote the same percent of themes with nPVIs in this range. These trends are almost identical when the theme cutoff is raised to 100 themes, which excludes Wagner and Strauss, Jr. (data not shown). Thus, very low nPVI themes are seemingly being replaced with very high nPVI themes while the middle range stays constant. One does not see these same trends when this analysis is performed on individual Italian composers (a culture that has been shown to maintain a constant nPVI over time). For Italians, all slopes in these ranges are below (or equal to) 0.03 and none of the regressions are significant (data not shown).

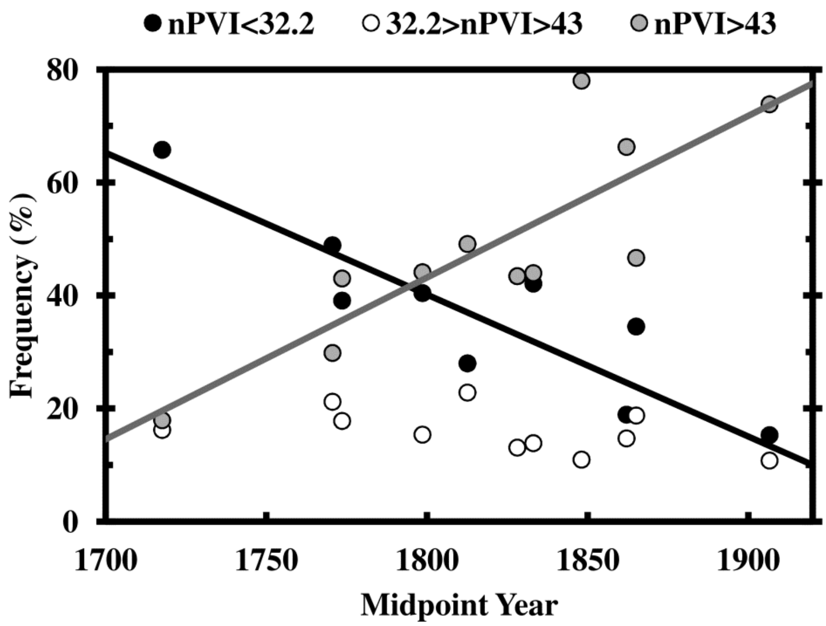


Figure 1. Percent of themes within a particular range for each composer plotted against midpoint year (see Table 1). Three vertical dots (black, grey, and white), representing the percent of themes within a particular nPVI range for each composer are plotted against midpoint year. Linear regressions were computed from the proportion of themes, for each composer, within each nPVI range. Lines represent significant linear regressions (p<0.01). Minimum number of themes per composer = 75.

While the Italian data gave important historical evidence for the potential uniqueness of this rhythmic shift in German/Austrian music, it was important to test what these nPVI proportions might look like if all composer nPVIs increased incrementally with time. A computer simulation was performed to test this hypothesis and the results are shown in **Figure 2**. Similar to the calculated composer data (**Figure 1**) the proportion of 'low nPVI' themes (>32.1) decreases in a significant manner (Frequency (%) = -0.22*(Midpoint Yr) +432.55, R^2=0.71, p<0.01) though the rate at is noticeably slower than the observed data (compare to slope 0.25, black line from **Figure 1**). The proportion of "high nPVI" themes (>43) also increases steadily and significantly (Frequency (%) = 0.26*(Midpoint Yr) – 423.53, R^2=0.70, p<0.01) yet this is also at a slower rate than the observed data (compare to slope 0.29, grey line from **Figure 1**). Finally, in stark contrast to the observed composer data (**Figure 1**) the proportion of "mid-range nPVI" themes (32.2 > nPVI > 43) in the model decreases significantly with time (Frequency (%) = -0.04*(Midpoint Yr) + 90.98, R^2=0.49, p=.017) while the observed composer data showed no significant change. Almost identical results were found when Wagner and Strauss, Jr. were excluded.

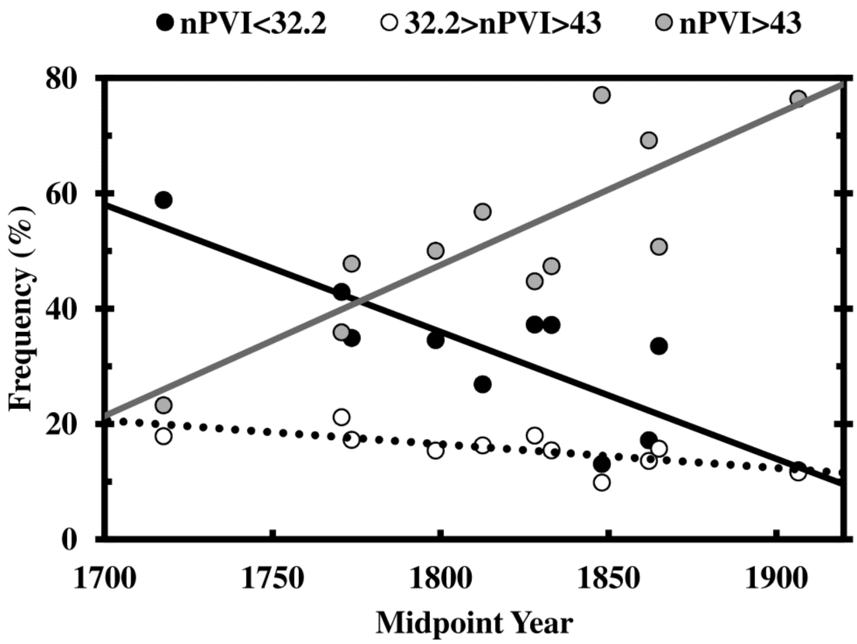


Figure 2. Predicted rhythmic changes over time using simulation software. Predicted percent of themes within a particular range for each composer (based on each composer's calculated mean and standard deviation) plotted against midpoint year. Three vertical dots (black, grey, and white), represent the percent of themes within a particular nPVI range for each composer and are plotted against midpoint year. Linear regressions were computed from the proportion of themes, for each composer, within an nPVI range. In contrast to Figure 1, all regression lines are significant; dotted line p=0.017, while solid lines are very significant (p≤0.01).

5. CONCLUSIONS

The present study illustrates how nPVI analysis can reveal the precise rhythmic underpinnings of important musical shifts. Thus, to understand which rhythms are changing (and when) the distribution of nPVI within each composer's corpus of themes was explored.

EVIDENCE FOR "THE REPLACEMENT HYPOTHESIS"

The suggestive increase in the mean nPVI of German/Austrian composers during the aforementioned 1760-1800 transition away from the Italian influence [1] prompted the development of a different method to study the "metric" of a composer's style, namely the distribution and spread of nPVI for all his/her themes (**Table 1 and Figure 1**). Testing the robustness and composition of this trend led to the observation that the frequency of themes with nPVI < 32.2 drops dramatically with time. These themes were replaced with a concomitant increase in themes with nPVI > 43. Themes in the middle range between 32.2 and 43 did not show any significant increase over time even when the theme cutoff was raised (**Figure 1**). A similar analysis of Italian composers (where one would expect no changes with time) revealed no significant linear regressions while a simulation of incremental nPVI increases with time also differed significantly from the observed results (**Figure 2**). These data offer suggestive evidence for the idea that 'low nPVI' themes in German music are being replaced with 'high nPVI' themes while the middle nPVI range stays constant ("the Replacement Hypothesis"). This is in contrast to an alternative hypothesis that the majority of nPVIs are modestly increasing to eventually resemble "more German" nPVIs by the mid- to late 1800s (e.g. **Figure 2**). More broadly, these data suggest what has been reported by musicologists (e.g. [10]) that composers in the late 1700's made a conscious choice to write "German" themes in place of the "Italianate" themes they once wrote.

SUGGESTIVE EVIDENCE FOR INCIPIENT 20TH CENTURY RHYTHMS

While most of the German/Austrian composers had a proportion of 'high nPVI' themes that ranged from 40-50% of themes, two composer's Richard Wagner and Richard Strauss, had nearly 80% of their themes in this category. With such strikingly different distributions, it is interesting to speculate that these composers' themes might be reflecting rhythms that are characteristic of 20th century composers, rather than the Romantic period they lived in. In fact, their style was most likely characteristic of the "New German School"; composers which pitted Beethoven's "absolute music" against their invention, the "symphonic poem". Importantly, these composers were not constrained by conventional musical forms e.g. sonata, concerto, etc. ([16], pp. 338-367). This break from the traditional compositional style can be seen as early as Wagner's *Tristan and Isolde* which is thought to have inspired future musical conventions like atonality, often used in "20th century" music ([17], p. 114). Future studies including "New German School" composers like Franz Liszt and 20th century German composers like Arnold Schoenberg could test if this predilection for very high nPVI themes is prevalent in these artists.

In conclusion, the current paper demonstrates the significance of using nPVI analysis as a tool to probe and quantify the precise rhythmic components of stylistic trends in music history. Thus, it provides a quantitative mechanism for comparison and support to the historical record and theory. Additional research using nPVI analysis could include the exploration of not only how complex stylistic trends in music history progress and develop, but also what the specific, underlying rhythmic structure of these trends might be.

REFERENCES

[1] J. R. Daniele and A. D. Patel, "An Empirical Study of Historical Patterns in Musical Rhythm," *Music Percept. Interdiscip. J.*, vol. 31, no. 1, pp. 10–18, Sep. 2013.

[2] A. D. Patel and J. R. Daniele, "An empirical comparison of rhythm in language and music," *Cognition*, vol. 87, no. 1, pp. B35–B45, Feb. 2003.

[3] E. ~L L. E Grabe, "Durational variability in speech and the rhythm class hypothesis," *Pap. Lab. Phonol.*, vol. Vol. 7, pp. 515–546, 2002.

[4] L. E. Ling, E. Grabe, and F. Nolan, "Quantitative characterizations of speech rhythm: syllable-timing in Singapore English," *Lang. Speech*, vol. 43, no. Pt 4, pp. 377–401, Dec. 2000.

[5] F. Ramus, "Acoustic correlates of linguistic rhythm: Perspectives," in *In: Proc. Speech Prosody 2002, Aix-en-Provence*.

[6] A. Arvaniti, "The usefulness of metrics in the quantification of speech rhythm," *J. Phon.*, vol. 40, no. 3, pp. 351–373, May 2012.

[7] J. R. Daniele and A. D. Patel, "Stability and Change in Rhythmic Patterning Across a Composer's Lifetime," *Music Percept. Interdiscip. J.*, vol. 33, no. 2, pp. 255–265, Dec. 2015.

[8] J. R. Daniele and A. D. Patel, *The Interplay of Linguistic and Historical Influences on Musical Rhythm in Different Cultures*. 2004.

[9] A. D. Patel and J. R. Daniele, "Stress-Timed vs. Syllable-Timed Music? A Comment on Huron and Ollen (2003)," *Music Percept. Interdiscip. J.*, vol. 21, no. 2, pp. 273–276, Dec. 2003.

[10] M. S. Morrow, *German music criticism in the late eighteenth century: aesthetic issues in instrumental music*. Cambridge, U.K.; New York: Cambridge University Press, 1997.

[11] N. C. Hansen, M. Sadakata, and M. Pearce, "Non-linear changes in the rhythm of European art music: Quantitative support for historical musicology.," *Music Percept. Interdiscip. J.*, 2016.

[12] H. Barlow and S. Morgenstern, *A Dictionary of musical themes.* London: Faber, 1983.

[13] S. Sadie and J. Tyrrell, *The new Grove dictionary of music and musicians*. London; New York: Distributed by Macmillan ; Distributed by Grove's Dictionaries, 2001.

[14] M. Solomon, *Beethoven*. New York; London: Schirmer Books ; Prentice Hall International, 1998.

[15] V. d'Indy, *Beethoven; a critical biography*. New York: Da Capo Press, 1970.

[16] A. Walker, *Franz Liszt. 2 2*. Ithaca, NY: Cornell Univ. Pr., 1993.

[17] J. Deathridge, *Wagner beyond good and evil*. Berkeley: University of California Press, 2008.

Perceptual-Motor Entrainment in the Context of the Speech-to-Song Illusion: Examining Differences in Tapping Patterns to Perceived Speech vs. Song

Alexandria Pabst[1], Christina Vanden Bosch der Nederlanden[2], Auriel Washburn[3], Drew Abney[1], Eric M. Chiu[1], and Ramesh Balasubramaniam[1]
[1]Cognitive and Information Sciences, University of California, Merced, Merced, CA, USA
[2]Department of Psychology, University of Nevada, Las Vegas, Las Vegas, NV, USA
[3]Department of Psychology, University of Cincinnati, Cincinnati, OH, USA

Language and music both require the listener to extract temporal information from utterances as they unfold in time. Although individuals display different patterns of coordination when synchronizing with music relative to speech, it is not clear whether these differences are driven by the way listeners entrain to music compared to language or by the acoustic features unique to each domain. In the current study, we used several ambiguous speech utterances that, when repeated, are perceived by both musicians and non-musicians as transforming from speech to song. These excerpts provide the opportunity to further understand how a listener's perception influences physical motor entrainment to temporal characteristics of excerpts perceived as speech or song. Listeners tapped to 10 repetitions of ambiguous, speech-to-song and unambiguous speech excerpts (synchronization) and continued tapping in silence after the repetitions (continuation). Following this task, participants were also asked to simply listen to three repetitions of each excerpt, and rate how closely they resembled speech or song. Preliminary results indicate that listeners have a higher degree of local entrainment to properties of ambiguous stimuli compared to unambiguous speech stimuli. Further analyses will examine the differences in multiscale sensorimotor coupling and whether these differences are driven by acoustic features or by top-down factors. We discuss our results in terms of the differences and similarities between language and music under the framework of multiscale sensorimotor coupling.

Please Don't Stop the Music: Song Completion in Patients with Aphasia

Anna Kasdan[1] and Swathi Kiran[2]
[1]Undergraduate Program in Neuroscience, Boston University, Boston, MA, USA
[2]Sargent College of Health and Rehabilitation Sciences, Boston University, Boston, MA, USA

Aphasia, an acquired language disorder resulting from brain damage, affects over one million individuals in the United States alone. Many persons with aphasia (PWA) have been observed to be able to sing the lyrics of songs more easily than they can speak the same words. The observation that not only singing, but even humming a melody, can facilitate speech output in PWA provided the foundation for Melodic Intonation Therapy. The current study aimed to look at PWA in their ability to complete phrases from songs in an experimental stem-completion format. Twenty PWA and 20 age-matched healthy controls participated. The task consisted of three conditions (sung, spoken, and melodic) each consisting of 20 well-known songs. Participants heard the first half of a phrase that was either sung, spoken, or intoned on the syllable "bum," and were asked to complete the phrase according to the format in which the stimulus was presented. Participants were scored on their ability to complete the melody and lyrics together in the sung condition, only the lyrics in the spoken condition, and only the melody in the melodic condition. PWA scored highest in the sung condition, followed by the spoken and melodic conditions, while controls scored comparably in the sung and spoken condition and much lower in the melodic condition. Both groups were better able to access the melody of songs in the sung condition than in the melodic only condition, while there was no difference in accuracy for lyric production between the sung and spoken conditions. These results are consistent with the integration hypothesis, which postulates that the text and tune of a song are integrated in memory. Moreover, there exists a stronger salience of the text over the tunes of songs in memory; this is true for PWA and controls. Interestingly, the more severe PWA scored higher in the melodic condition than in the spoken condition, while the opposite trend was found for less severe PWA and controls. This indicates that access to melody is preserved in PWA, particularly those who are more severe, even while they exhibit language impairments. Findings may have implications for using music as a more widely implemented tool in speech therapy for PWA.

Do musical abilities predict reading, spelling, and mathematical abilities in young adults?

Sarah R. Hansen[1] and Kathleen A. Corrigall[1]
[1]Department of Psychology, MacEwan University, Edmonton, Alberta, Canada

Mounting evidence suggests that musical abilities are closely associated with reading ability, particularly word decoding ("sounding out" or pronouncing individual words). For example, musically trained children typically show superior reading abilities compared to untrained children, and music perception skills are associated reading skills. Remaining questions include 1) whether associations remain after controlling for potential confounding variables (e.g., demographic, cognitive, and personality variables), 2) whether the associations are specific to word decoding or extend to other academic skills (e.g., reading comprehension, spelling skills, and mathematical ability), 3) which aspects of music perception best predict reading skills, and 4) whether the associations are also evident in adult samples. In the current study, we examined the specificity of the association between musical abilities and reading skills in undergraduate students. Participants completed measures of academic skills (word decoding, reading comprehension, spelling, and math; Wide Range Achievement Test), personality (Big Five Inventory), IQ (Wechsler Abbreviated Scale of Intelligence – Short Form), and music perception (Mini Profile of Music Perception Skills; Beat Alignment Test), and they provided musical background and demographic information (e.g., age, gender, parents' education, family income). Results (n = 105) suggest that music perception skills are associated with spelling and word decoding but not reading comprehension or mathematical abilities. Associations between music perception skills and word decoding were especially strong and remained significant even when potential confounds (e.g., IQ) were statistically controlled. Our findings provide further support that musical abilities are specifically associated with reading and reading-related skills (e.g., spelling) that rely most heavily on sound processing, even into adulthood.

Investigating music schemas as a form of linguistic constructions in the music of the Dong people in Guizhou, China

Kathryn Davies[1]
[1]Department of Music, University of Hong Kong, Hong Kong SAR, China

Research on musical schemas has revealed stock musical patterns that serve particular functions within a musical piece. For example, Gjerdingen (2007) has shown that this system of musical schemas was a common means of musical production in the Western classical music tradition. In a similar manner, within linguistics, construction grammar researchers argue that language utilizes multiple levels of form-function pairings: like the semiotic pairing of a form, sound, and meaning to form a word, even sentences carry an arbitrary, but conventionalized, pairing of a certain structure or form and a meaning (e.g. *X causes Y to Z*, *the Xer the Yer*). There are several parallels between musical schema and linguistic constructions (cf. Gjerdingen & Bourne 2015). For instance, both entail a system of categories, from which certain elements can be selected and interchanged in the various “slots” in the forms. Nevertheless, to draw conclusions about common cognitive mechanisms for language and music, further, cross-cultural evidence of musical schemas is needed.

I will present exploratory findings of research on the songs of the ethnic minority Dong people in Guizhou, China. Historically, the Dong people have no orthography for their language (*Kam*), yet they have a rich tradition of singing (cf. Long & Zheng 1998; Ingram 2014). This research will investigate if and how they may use a similar form-function pairing of musical chunks. For example, their use of a *sor* or “musical habitus” (Ingram 2012) may be indications of categorization of musical entities and prefabricated chunking similar to constructions in language. Another indicator is their rising, second interval musical “idiom” that is often used to signal the end of musical sections or songs. The Dong singing tradition offers an opportunity to investigate a human cultural artefact to help understand the cognitive relationship between music and language. Moreover, without an orthographic system, the Dong people may use music as a cultural replacement of some aspects of language.

Characteristics of Non-linguistic Vocalizations as Auditory Emoticons

Kyoungsoo Chun[1], Jinah Kwak[1] and Juhan Nam[1]
[1]Graduate School of Culture Technology, KAIST, Daejeon, South Korea

Non-verbal communication is important part of interpersonal communication. Among others, facial expressions and non-linguistic vocalizations (e.g. laughing or crying) are used as effective means to deliver messages without words or control the strength of messages in words particularly when emotions are involved. These days, in computer-mediated communication (CMC), emoticons have played such a role in a non-verbal manner. While the text or image-based symbols have been popularly used, their acoustic counterpart, or non-linguistic vocalizations have been rarely handled in the CMC. In this paper, we explore the possibility of using non-linguistic vocalization sounds as an auditory emotion. To this end, we built a dataset of non-linguistic vocalization sounds and surveyed how the sounds are emotionally well-paired with emoticons and what characteristics they have in semantic and acoustic aspects. Specifically, we focused on cuteness, naturalness, childness and gender in semantic aspect, considering general properties of image and text emoticons. In acoustic aspect, we analyzed pitch, energy, onset and length. The results show that, in determining the suitability as an auditory emoticon, different characteristics are important depending on emotion types. However, those recognized as cute and natural, and with relatively high pitch or short length, are generally more appropriate as an auditory emoticon. This semantic and acoustic analysis may be used as a guide for designing auditory emoticons.

Categorization in the speech to song transformation (STS)

Bas Cornelissen[1], Makiko Sadakata[1,2] and Henkjan Honing[1,2]
[1]Music Cognition Group, Institute for Logic, Language and Computation, University of Amsterdam, The Netherlands
[2]Artificial Intelligence department, Radboud University Nijmegen, The Netherlands

Repetition of speech has long been known to cause perceptual transformations. In the semantic satiation and verbal transformation effect, repetition for example leads to loss and shifts of meaning. Similarly, in the speech to song transformation (STS) a spoken phrase starts to sound like a song when repeated several times, even though the phrase itself is not changed (Deutsch et al., 2011). This makes it a powerful tool for investigating the differences between speech and song perception without introducing confounding acoustical differences. While speech repetition triggers various transformations, it remains an open question why certain phrases transform to song and others do not. Several studies have touched upon this question and we aim to combine these to provide a general answer by construing the STS as a categorization problem. Previous studies suggest that pitch and rhythm related features strongly contribute to the transformation (e.g. Deutsch et al., 2011; Falk et al., 2014; Tierney et al., 2012). The most detailed study of such features (Falk et al. 2014) investigated only a small set of stimuli, whereas Tierney et al. (2012) analyzed many phrases, but used simple pitch and timing measures. The current study combines both approaches using a web-based experiment with nearly 300 spoken fragments. We analyze the ratings using different measures, including paired variability index and intensity weighted standard deviations. The result provides additional support for the previous findings and suggest that the extent to which a phrase can be heard as having a clear and musical pitch structure, predicts the STS. In summary, these findings strengthen the idea that speech phrases that transform to song have clear song-like features to start with. This highlights the problem of inferring the transformed melody from the sound, which suggests a computational model of the STS. Moreover, it has implications on perceptual shifts in repeated speech more generally.

Temporal Event Clustering differs in Infant- versus Adult-Directed Speech and Song

Christopher T. Kello[1], Simone Falk[2]
[1] Cognitive and Information Sciences, University of California, Merced, USA
[2] Department of German Philology, LMU, Munich, Germany

Adults speak and sing differently when directing their vocalizations to infants and young children compared with other adults. Prior studies have shown that infant-directed vocalizations tend to have higher fundamental frequencies, slower tempos, and greater variations in tonal contours. Changes in the acoustic qualities of infant-directed vocalizations appear to affect infants—attracting their attention, raising arousal levels, and even facilitating speech perception—but it is not clear what properties of infant-directed vocalizations are most relevant to infant speech and song perception. In the present study, we investigate the temporal dynamics of infant vs. adult-directed speech and song using a new methodology designed to analyze temporal clustering in acoustic events. We hypothesize that temporal clustering in spoken language is relevant to infant speech and song perception because it reflects the hierarchical structure of spoken language, by virtue of greater temporal clustering in infant-directed vocalizations. Acoustic events are peaks in the acoustic amplitudes of an auditory filter bank that spans a range of frequencies, and temporal clustering is measured with Allan Factor analyses used previously to measure temporal clustering in neural spike trains. Fifteen German mothers read a story and sang a nursery rhyme for their 6 months old infants, and also for an adult. Analyses showed greater temporal clustering in infant-directed compared to adult-directed versions of story-reading and singing, particularly at longer timescales on the order of seconds. These results open new ways of analyzing the temporal dynamics of speech and song, and relating them to infant language acquisition and musical development.

Relation of Musical Training to Vocoded Speech Recognition

Tahnee Marquardt[1], Sara Prostko[1], Gina Scharenbroch[1,2], and Terry L. Gottfried[1]
[1]Department of Psychology, Lawrence University, Appleton, WI, USA
[2]Department of Speech-Language-Hearing Sciences, University of Minnesota, Minneapolis, MN, USA

Research has shown that aspects of musical training and ability transfer to speech perception in an unfamiliar language (e.g., Mandarin tones, Gottfried, 2007) and in one's native language (e.g., enhanced speech encoding in noise, Zendel et al., 2015). These findings led us to test whether a musician advantage might also be observed in comprehension of acoustically degraded speech. Normal-hearing listeners performed a pretest in which they transcribed 8-channel TigerCIS vocoded speech (which presents information similar to that provided by cochlear implants; see Loebach et al., 2008). These listeners were then trained (with feedback) on 60 vocoded Harvard sentences and 40 MRT (Modified Rhyme Test) words or the same sentences and words in natural speech. Musicians (conservatory-trained) and non-musicians were compared on their initial accuracy in transcribing the vocoded speech, and their transcription accuracy after the limited amount of training on natural or vocoded speech. Accuracy was significantly greater on sentences than words. However, musicians and non-musicians were not significantly different in their transcription accuracy, neither on the pretest nor on tests after training (new 8-channel and 4-channel stimuli). Accuracy for all listeners on 8-channel posttests was greater than on pretests, but there was no significant difference in posttest performance between those trained on natural vs. vocoded speech. Further tests using a more extensive training regimen and perceptual tests to assess listeners' musical pitch and rhythm abilities are currently being conducted to determine whether specific musical skills are more relevant to improved speech perception.

High Talkers, Low Talkers, Yada, Yada, Yada: What Seinfeld tells us about emotional expression in speech and music

Tiffany Bamdad[1], Brooke M. Okada[1], and L. Robert Slevc[1]
[1] Department of Psychology, University of Maryland, College Park, MD, USA

While both music and speech can express emotion, it is not yet clear if they do so via a "common code." On one hand, speech prosody and music convey emotion in similar ways—through manipulation of rate, intensity, and pitch. On the other hand, relationships between discrete pitches are critical to the emotional character of musical pieces, whereas speech prosody does not seem to rely on discrete pitch. Supporting a common code for emotion in music and speech, Curtis and Bharucha (2010) found that pitches in sad speech approximate a minor third (an interval associated with sad music). This may even occur *between* speakers: Okada et al. (2012) found that individuals tend to speak in keys separated by dissonant intervals when disagreeing, and by consonant intervals when agreeing. However, this conclusion required extracting a likely key from spoken phrases, thus relied on approximations of both pitch and key. The current study more directly tests whether intervals between interlocutors' spoken pitches relate to the quality of their conversation. We selected agreeable and disagreeable conversations from *Seinfeld* episodes and compared the pitch of the last word of one character's utterance with the pitch of the first word of their conversation partner's response. Distance between the pitches was approximated to the closest musical interval and categorized as a perfect consonant, imperfect consonant, or dissonant interval. In contrast to previous findings, agreeable conversations did not contain more perfect or imperfect consonant intervals compared to disagreeable conversations, and disagreeable conversations did not contain more dissonant or imperfect consonant intervals compared to agreeable conversations. These results suggest that expression of emotion in speech does not manifest in musically relevant intervals between interlocutors, supporting the view that emotion in speech and music does not rely on a common code, at least between speakers in conversation.

New, freely-available technology for studying the impact of simple musical training on speech perception in cochlear implant users

Thomas R. Colgrove,[*1] Aniruddh D. Patel, [#2]

[*] *Departments of Computer Science and Music (BS 2015), Tufts University, USA*

[#] *Department of Psychology, Tufts University, USA*

[1]trcolgrove@gmail.com, [2]a.patel@tufts.edu

ABSTRACT

Cochlear implants (CIs) provide good understanding of speech under ideal conditions (e.g., in quiet), but most CI users have difficulty perceiving pitch patterns accurately. Good pitch pattern perception is essential for the perception of musical melody and speech intonation. Current CI technology does not provide sufficient detail for accurate pitch processing. However, auditory training has been shown to benefit CI users' speech perception, raising the idea that training could also be used to enhance pitch pattern perception. We wish to determine if musical instrument training (learning to play simple pitch patterns on a piano keyboard) enhances the perception of pitch patterns in music and speech. We hypothesize that auditory-motor keyboard training will be more effective at improving pitch pattern perception than purely auditory training, because playing pitch patterns on a musical instrument creates an action-perception link between the motor and auditory systems. To this end, we present a new software platform called 'Contours', which provides simple musical keyboard training in a format aimed at individuals with no prior musical experience. The software runs on an Android tablet connected to a portable 3-octave MIDI keyboard and employs a visual representation of melodic contour shapes that aims to eliminate the steep learning curve of traditional music notation. Additionally, the software records the user's accuracy in playing the contours, provides feedback to the user throughout the training, and saves performance data to a web server for data analysis. The length and difficulty of contours can be gradually increased over the course of training, and several different sounds can be used, from simple sine waves to piano tones. By testing auditory perception before and after training, one can see if training using Contours enhances the perception of pitch patterns, and of music and speech more generally.

I. INTRODUCTION

Although cochlear implants provide good understanding of spoken words under ideal conditions (e.g., speech in quiet, familiar voices), most cochlear implant (CI) users have great difficulty perceiving pitch patterns accurately (Limb and Roy 2014). Good pitch pattern perception is essential to music appreciation and to understanding the "melody of speech" (the ups and downs of voice pitch, which help signal phrase boundaries and emotion). Current CI technology does not provide sufficient detail for accurate pitch processing. Much research and development has been directed at improving the spectro-temporal resolution of the CI (e.g., "virtual channels" between the implanted electrodes, high-rate stimulation), but with little-to-no improvement over standard implant technology. However, auditory training has been shown to benefit CI users' speech perception. This raises the idea that training could also be used to enhance pitch pattern perception.

In fact, research has shown that purely perceptual training can improve pitch pattern perception in CI users (Galvin et al., 2009). This training involves matching simple pitch contours to their visual analogs, based on Galvin et al.'s (2007) "Melodic Contour Identification" (MCI) test. In this test, a listener hears one of 9 pitch contour shapes (e.g., rising, falling, rising-falling) and identifies it by clicking on the corresponding image on a computer screen (see Figure 1 below). This is an easy task for normal listeners but a very difficult task for CI users, due to their problems with pitch pattern perception. Galvin et al. (2009) showed that by training on the MCI task (using different root notes than used for testing), CI users can improve their performance on the MCI test.

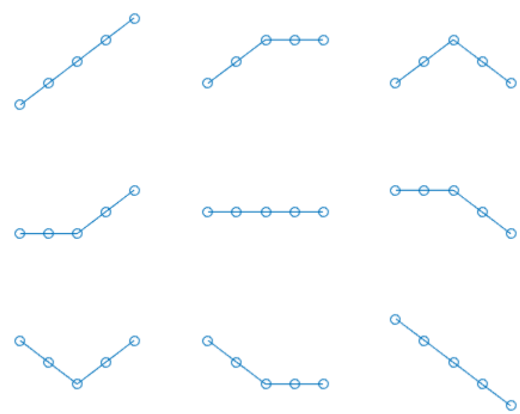


Figure 1. Schematic of the nine pitch contours in the Melodic Contour Identification (MCI) test (Galvin et al., 2007). After hearing a pitch pattern, the listener identifies the pattern by choosing the corresponding contour on a computer screen (e.g., rising, falling, rising-falling etc.). Each pattern has 5 pitches, and the pitch distance between notes in a contour can be systematically varied from small to large. These serve as the easy contours in the Contours program.

The current research is motivated by the idea that CI users can enhance their pitch contour perception by training that involves **playing** pitch contours on a piano keyboard. We hypothesize that keyboard training will prove more effective at improving pitch pattern perception than purely auditory training. This is because playing pitch patterns on a musical instrument (vs. simply hearing them) creates an action-perception link between the motor and auditory systems of the brain, which is absent in purely perceptual training (Lahav et al. 2007; Chen et al., 2012). There is evidence from auditory neuroscience that training involving action-perception coupling leads to more effective neural changes in sound processing than does training based on listening alone (e.g., Lappé et al., 2008; Matthias et al., 2015). This empirical work provides the background and rationale for our hypothesis.

The feasibility of this type of training has been demonstrated in a pilot study in which two CI users underwent this type of training (see Patel, 2014 for details). Training consisted of playing patterns based on the pitch contours in Figure 1, using either 5 successive white keys (1-2 semitone spacing) or black keys (2-3 semitone spacing) on a piano keyboard. These spacings were deemed optimal as many CI users have difficulty perceiving pitch differences less than 2 semitones. Thus the music training aimed to develop greater precision in pitch contour processing than the CI users normally have, using a form of sensorimotor training in which the participants themselves produce the auditory pattern they are learning. After keyboard training, both participants improved on the MCI test. One participant also showed an improvement in speech-in-noise perception, while the other showed an improvement in statement-question identification based on linguistic pitch contour (data reported in Patel, 2014).

To facilitate further research of this type, we have developed a new software program ('Contours') which interfaces with a MIDI keyboard and trains a user to play melodic contours of different shapes, while tracking progress by logging performance data. No prior musical experience is assumed of the user. This software is considerably more flexible than the software used in the study of the two CI users reported in Patel (2014). For example, the new software offers three different levels of contour complexity, allowing the user to progress from simpler to more challenging contours, which should help keep the training interesting and motivate participants to practice. Easy contours are the contour shapes in Figure 1: these have 5 notes and at most one direction change in the contour (e.g., from rising to falling, or rising to flat). Medium difficulty contours are shown in Figure 2: these have 6 notes and 2 contour direction changes. Hard difficulty contours are shown in figure 3: these have 7 notes and 3 contour direction changes. At each level of difficulty, the program presents each contour shape at several different transposition levels, allowing the user to play the contour across a range of frequencies.

Another aspect of the flexibility of Contours is that at each difficulty level, the user has options for the size of pitch gaps between notes in the contours. For CI users, different pitch contours shapes are considerably more challenging to discern when the gaps between notes are small vs. large (see Galvin et al., 2009, Figure 3). In Contours, pitch gaps can be set so that the gap size between pitches is 0, 1, or 2 white notes on the keyboard. (To simplify the playing of the piano keyboard for non-musicians, black keys are not used in during the training.)

One final way in which Contours is flexible is that the user is given 5 choices for the sound produced by the keyboard: sine-waves (the easiest sounds for CI users to perceive), square waves, triangle waves, sawtooth waves, and sampled piano tones. Thus users can learn what a pitch contour sounds like when played in different timbres, which will hopefully allow generalization and recognition of contours when created by different sound sources (including the human voice).

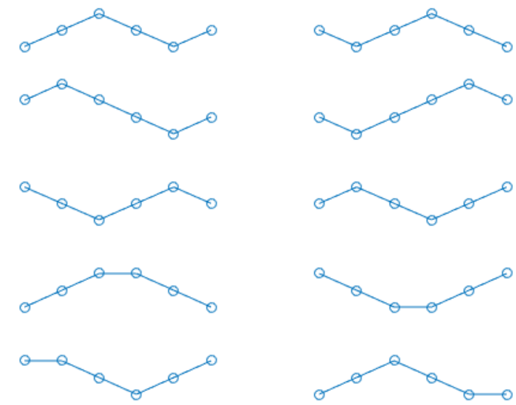


Figure 2. Schematic of the 10 pitch contours in the medium difficulty condition of the Contours program. Each pattern has 6 pitches and 2 changes of contour direction (e.g., rising to falling, level to falling). The pitch distance between notes in a contour can be systematically varied from small to large.

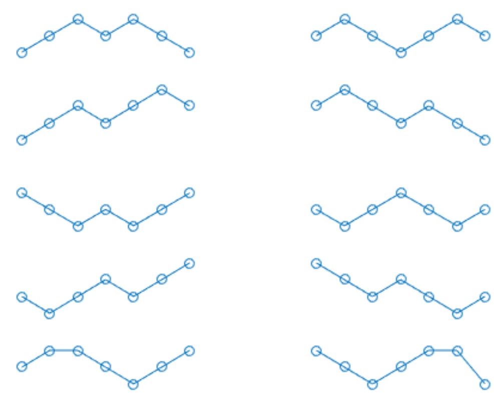


Figure 3. Schematic of the 10 pitch contours in the hard difficulty condition of the Contours program. Each pattern has 7 pitches and 3 changes of contour direction (e.g., rising to falling, level to falling). The pitch distance between notes in a contour can be systematically varied from small to large.

Apart from the flexibility of Contours, another other important feature of the program is data logging, which records the timing and accuracy with which each contour is played, allowing researchers to quantify the improvement in contour performance over time (for details, see the next section: unit testing revealed that logged times have an accuracy of +/- 100 ms). Importantly, Contours has a game-like visual interface, which presents contours as visual shapes using attractive graphics, dynamically shows the user which note to play next in a contour, and gives visual feedback each time a mistake is made or a contour is played correctly.

Contours is freely available and open-source, allowing researchers to customize it to their needs, e.g., by adding other contours shapes, sounds, or gaps sizes between notes.

II. DETAILS OF TRAINING

A.Notation

An important feature of Contours is the simplicity and readability of the musical notation. While traditional music notation has a steep learning curve, the notation in Contours uses color as a visual aid to assist users in playing the patterns presented (Figure 4), and does not require learning note names (e.g., C, D, E, etc.) The notation uses unstemmed circles to represent pitches. Each circle has a specific color, which corresponds to a color on the MIDI keyboard (Figure 5). The notation and keyboard span a 3 octave range (from C2 to C5, i.e., MIDI note 48 to 84), with the pattern of colors repeating each octave (e.g., all Cs on the keyboard are purple, across different octaves).

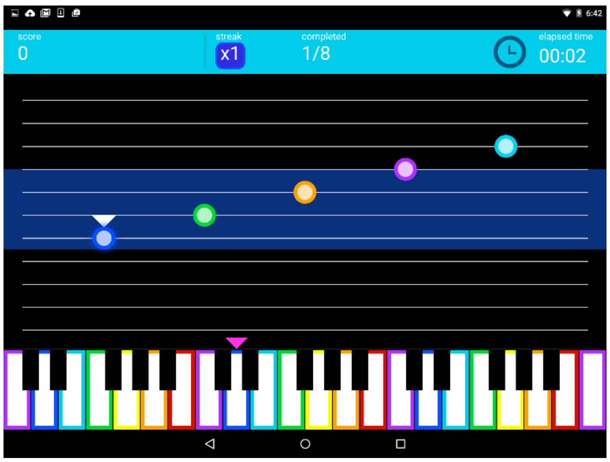


Figure 4. Screenshot of Contours, presenting an easy contour to the user. The next note to be played is indicated by an inverted wedge over the circle in the music notation, and over the key that should be pressed on the keyboard to produce that note. When the corresponding key is pressed, the wedge moves to the next pitch in the music notation and the corresponding key on the image of the keyboard. The image of the keyboard at the bottom of the Contours screen is for visual reference only: the user plays a physical keyboard attached to the tablet, which has the same color scheme (see Figure 5).

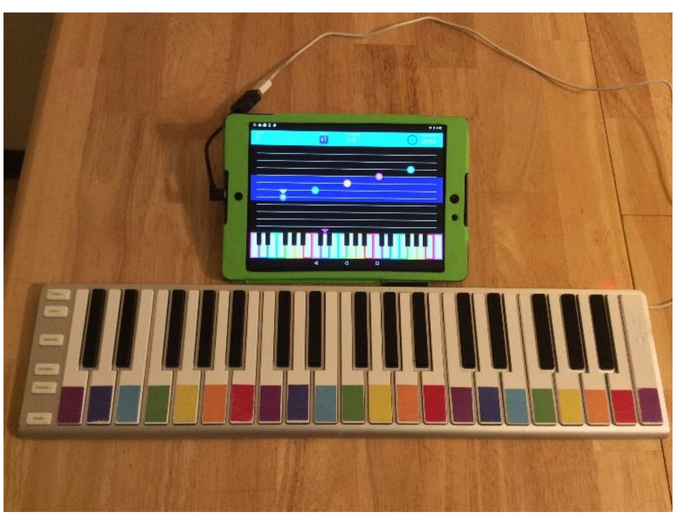


Figure 5. Contours running on an Android Tablet, connected to the color-coded 3-octave MIDI keyboard. Colors are placed on the white keys using adhesive paper (no black keys are used in training). Each note within an octave has a unique color, and the color pattern repeats across octaves.

As with Western music notation, the staff lines and spaces map in a one-to-one fashion to the large keys of the keyboard (i.e., the white keys, whose tips are covered with colors, cf. Figure 5). Thus the contour shown in Figure 4 involves playing an ascending contour which starts on a large blue key (near the middle of the keyboard) and skips one large key between each of its pitches. An important difference between the music notation used in Contours and Western music notation is that the notation in Contours does not include any accidentals. (Correspondingly, none of the black keys are used during training.) This is to maintain the simplicity of notation and make it easy to learn for individuals with no prior musical training.

B. Training Program Flow

i. Each time a user plays Contours, they select a contour difficulty (easy, medium or hard, cf. Figs 1-3), a pitch gap size between successive notes in the contours (narrow, medium, or wide), and a sound for the application to play (sine wave, triangle wave, square wave, sawtooth wave, or sampled piano sounds). For pitch gap size, narrow, medium and wide correspond to the number of colored keys skipped when playing consecutive notes in a contour: i.e., 0, 1, or 2, respectively. (Thus the ascending contour shown in Figure 4 has a medium gap size).
    i. There is no enforced order in which the user progresses through the different conditions. This allows the researcher suggest a particular order, e.g., easy level first, then medium, then hard, with wide pitch gaps at all levels, followed by repeating these three levels with medium pitch gaps, and then again with narrow pitch gaps, all using sine-wave tones. This example would result in 9 conditions (3 contour difficulty levels x 3 pitch gap sizes), all using the sine-wave sound. Repeating this regimen with another sound (such as the square wave) would add 9 more conditions.

ii. Once the user chooses a difficulty level, a pitch gap size, and a sound, Contours presents the user with a block of contours consisting of each contour shape from that difficulty level at 5 different transpositions. (Hence a block has 45 contours at the easy level, and 50 at the medium and hard levels, cf. Figs 1-3.) For a given contour shape, transpositions are chosen randomly (without replacement) from all possible transpositions that fit within the 3 octave range. Thus, 5 different transpositions are always presented, and repeating a block will result in a new set of randomly-chosen transpositions.

iii. When playing a given contour, if the user makes a mistake (e.g., plays a wrong note), a visual message appears in the center of the screen (e.g., “Whoops!”) and the user must begin the contour again from the first note. Upon playing a contour correctly, the user gets a visual congratulation message (e.g., “Good job!”) and earns points, which appear in the top bar of the screen. (See section on Data metrics for a description of how points are calculated.)

iv. Once a block of contours is completed, the user is asked a few short questions about the contours played, and responds to each question using a number from 1 to 5. The questions are:
   i. How difficult was the task?
   ii. How different did the contours sound from each other?
   iii. Do you think you are improving at the task?

v. Once a block is completed, quantitative data about the user's performance on each of the 45 (or 50) contours is logged, such as number of errors, speed and accuracy. These data are stored locally on the Android device, and sent to a remote server in JSON format (see section on Data metrics for details). These data can be downloaded from the server as spreadsheets in .csv format for analysis. This allows researchers to quantify the progress of users over multiple sessions. (Responses to the questions described above are also stored.)

C. Data Metrics

Within a block, the user earns points for each correctly-played contour. Points reflect both speed and accuracy of performance. The score for a contour = 100 points (for completing the contour correctly), plus a bonus which equals 100 minus the total time it took to complete the contour (in milliseconds/250). This number is then multiplied by an integer which corresponds to how many contours have been completed in a row without any errors (the maximum multiplier is 8). Finally, this number is added to the existing score to create a running point total for the block. After completing a contour, the top bar briefly displays the points earned for that contour, and the running total for the block, with the latter score remaining on screen throughout the block.

When a block is completed, performance data about that block are stored locally on the tablet, and then uploaded to a server when an internet connection is available. These data can then be downloaded from the server as a .csv spreadsheet by the researcher for analysis, e.g., to quantify improvement over time. Each spreadsheet has 11 columns. The columns are as follows:

1. Contour ID: an integer between 1 and 9 (for the easy contour condition) or 1 and 10 (for the medium and hard conditions), specifying contour shape. For the easy condition, the numbers correspond to the layout of shapes in Figure 1, in the following order:
1 2 3
4 5 6
7 8 9
For the medium and hard conditions, the numbers correspond to the layout of shapes in Figure 2 and 3, respectively, in the following order:
1 2
3 4
5 6
7 8
9 10
2. Difficulty: Easy, medium or hard (cf. Figs 1 – 3)
3.Note gap: an integer between 0 and 2, indicating how many colored keys are skipped between consecutive notes of the contour.
4.**Sound**: specifies which of the 5 sounds was used: sine wave, square wave, triangle wave, sawtooth wave, sampled piano.
5. Start note: The MIDI note number of the first note of the contour (indicates the transposition level).
6. Total completion time: Time from when the contour is first posted on the screen to when it is completed correctly. If mistakes were made, and the user had to start the contour again, this time is included. (However, time taken by the posting of messages triggered by errors, e.g., "Whoops!", is not included.)
7. Num errors: The number of wrong notes played before completing the contour correctly.
8. Percent error: Num errors/number of notes played while completing a contour.
9. Completion time (correct run): Time between the onset of the first note and the onset of the last note when the contour is played correctly
10. Note interonset interval sd: The standard deviation of durations between note onsets when the contour is played correctly.
11. Date and time
Date and time the contour was completed.

Details on how to access data spreadsheets from a server are provided on the GitHub repository for the Contours project.

III. TECHNICAL SPECIFICS AND EXTENSIBLITY

The Contours training application for Android can be downloaded from GitHub and installed on any Android tablet device, although tablets with at least 9" of screen space, running Android 5.0 or higher are recommended. This tablet should be connected to a 3-octave MIDI keyboard to which colored adhesive paper has been added to match the color scheme in the software (see Figure 5). We used a CME Xkey 37.

The application is open source under the MIT license (https://opensource.org/licenses/MIT), and can be freely used, shared, and modified to suit different testing needs. While there are many possible extensions of the program, there are a few that can be made with relatively minimal coding (see below). We chose Android as the operating system for its portability and affordability, the open nature of the platform, and the ability to use the application on many different devices.

A. Modification of Contour Shapes in XML

The preloaded musical contour shapes are defined in a project xml contours.xml file, that can be easily modified to provide other contours of different lengths and shapes. This allows researchers to easily expand on or replace the preloaded set of contours. For example, some users may want to progress beyond the preloaded contours to more challenging and complex shapes. Each contour shape only needs to be specified once, as the software will automatically transpose the contour to multiple different pitch levels. Example xml codings of three contours are shown below. Note that contours are specified using standard note names from Western music (e.g., C2, E2), which combine a pitch class and an octave number. (Further documentation on code specifics can be found on the project github.)

```
<!--Rising one skip -->
    <item>C2,E2,G2,B2,D3</item>
<!-- Rising Flat one skip -->
    <item>D2,F2,A2,A2,A2</item>
<!-- Rising Falling one skip -->
    <item>C2,E2,G2,E2,C2</item>
```

3 *xml formatted contours, 5 notes each*

B. Modify Application Timbre using Pure Data

One powerful component of Contours is the ability to test the effect of various timbral qualities on pitch perception and pitch contour learning among CI users. The synthesizer component of the application uses the Pure Data visual language to enable the generation of a wide array of different musical sounds. The integration of Pure Data with Contours is modular and allows modification and swapping of pure data patches. It is straightforward to modify or swap out the currently included Pure Data *patches*, to use an expanded set of sounds. The two currently included modules are a fully functional subtractive synth based on a related implementation written by Christopher Penny, and an instrument sampler. Possible modifications include modifying the existing subtractive synth, changing the sample sets available on the sampler, or to using a different patch altogether. More detailed instructions for extending the Contours synthesizer can be found in the project documentation.

C. Example Server

The Contours application comes preloaded with the capability to send detailed test results to a server application in JSON format. A fully-featured example server that can be used with the contours server is provided on the application GitHub. The server is coded in a Flask/Python/MongoDB back end. The example server also includes a simple front end webpage which displays summary test results in tabular form, and enables the spreadsheet download of results for each completed block of contours in CSV format. Administrators of the Contours training program may use the example server as provided, extend it, or use their own server.

D. Further Extensions

More extensive changes to the contours application require some knowledge of Java and the Android SDK. Brief descriptions of possible future extensions will be detailed in the following section.

IV. PROPOSALS FOR FUTURE CAPABILITIES

The Contours application currently tests the effects of simple musical keyboard training on CI users ability to perceive pitch contours. However, there are multiple ways in which this application could be significantly extended. For instance, the rhythm and cadence of speech is arguably as important as pitch variation. Furthermore, rhythm is often a fundamental component of people's ability to identify different melodies. It is possible that introducing a rhythmic component to the training program could be of greater benefit to CI users in understanding speech.

Furthermore the keyboard training at present only presents the user with monophonic melodic patterns. It would be fairly straightforward to extend the training to include contours featuring chords, or more than one note played at the same time.

V. CONCLUSION

Contours provides a flexible, sophisticated platform for training CI users to play simple melodic contours, with the goal of improving their pitch contour perception, with consequent benefits for both music and speech perception. The software assumes no prior musical knowledge or experience. Existing perceptual tests, such as the melodic contour identification (MCI) test (Galvin et al., 2007), can be given before and after training to determine if learning to play melodic contours enhances the ability to perceive and discriminate pitch contours. Ultimately, one could compare such training to purely auditory training (matched in duration) to test the hypothesis that actively playing contours is more effective for enhancing pitch perception, due to the auditory-motor nature of such training.

The software is available for download at https://github.com/contoursapp/Contours

ACKNOWLEDGMENT

This research was supported by a grant from the Paul S. Veneklasen Research Foundation. We thank Allison Reid for her help in developing the medium and hard difficulty contours and test the Contours software, and Paul Lehrman and Chris Penny for their technical insights.

REFERENCES

Galvin, J. J., Fu, Q. J., & Nogaki, G. (2007). Melodic contour identification by cochlear implant listeners. *Ear and hearing*, *28,* 302-319.

Galvin, J. J., Fu, Q. J., & Shannon, R. V. (2009). Melodic contour identification and music perception by cochlear implant users. *Annals of the New York Academy of Sciences*, *1169,* 518-533.

Lappé, C., Herholz, S. C., Trainor, L. J., & Pantev, C. (2008). Cortical plasticity induced by short-term unimodal and multimodal musical training. *The Journal of Neuroscience*, *28,* 9632-9639.

Lahav, A., Saltzman, E., & Schlaug, G. (2007). Action representation of sound: audiomotor recognition network while listening to newly acquired actions. *The Journal of Neuroscience, 27,* 308-314.

Limb C.J., & Roy A.T. (2014). Technological, biological, and acoustical constraints to music perception in cochlear implant users. *Hearing Research, 308,* 13-26.

Mathias, B., Palmer, C., Perrin, F., & Tillmann, B. (2015). Sensorimotor learning enhances expectations during auditory perception. *Cerebral Cortex*, *25,* 2238-2254.

Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing research*, *308,* 98-108.

Music and language: Syntactic interaction effects found without violations, and modulated by attention

Anna Fiveash[1], Genevieve McArthur[2] and William F. Thompson[1]
[1]Department of Psychology, Macquarie University, Sydney, NSW, Australia
[2]Department of Cognitive Science, Macquarie University, Sydney, NSW, Australia

Research suggests that music and language draw on similar syntactic processing resources, but questions remain about the nature of these resources, and the degree to which they are shared. Evidence suggests that performance is impaired when both domains simultaneously process syntax. However, most of this evidence comes from experiments that introduce syntactic errors in music (e.g. out-of-key notes) or language (e.g. grammatically incorrect sentences). Such experiments cannot rule out the possibility that effects are due to shared error detection mechanisms independent to syntactic processing. In a series of experiments, we examined whether evidence for shared syntactic resources by music and language can be elicited without introducing syntactic errors, and whether such interference effects are modulated by attention. In Experiment 1, participants listened to audio stimuli while reading language. We discovered that memory for complex sentences (with syntax), but not word-lists (no syntax), was (1) lower when paired with a single-timbre melody than when paired with environmental sounds, and (2) higher when paired with a melody with alternating timbres than when paired with a single-timbre melody. In Experiment 2, participants did a same/different task, which found that the latter effect occurred because alternating timbres drew attention away from syntax, thereby releasing resources for processing syntax. In Experiment 3, participants selectively attended to one stream of audio stimuli, while reading language. Results showed that memory for complex sentences was worse when participants attended to melodies while ignoring environmental sounds, and better when participants attended to environmental sounds while ignoring melodies. This suggests that attention modulates syntactic processing. Overall, the results indicate that syntactic interaction effects cannot be explained by overlapping error detection mechanisms, and that attention modulates syntactic processing.

Creatures of context: Recruiting musical or linguistic knowledge for spoken, sung, and ambiguous utterances

Christina Vanden Bosch der Nederlanden, Erin E. Hannon[1], and Joel S. Snyder[1]
[1]Department of Psychology, University of Nevada, Las Vegas; Las Vegas, NV, USA

Every day we use our experiences and knowledge to make sense of the sights and sounds around us. Although there is a wealth of literature examining speech and music processing independently, there is little known about how we flexibly recruit musical and linguistic knowledge in different contexts while also controlling for differences in acoustic features. As such, it remains unclear whether children are able to recruit domain-specific knowledge even when acoustic features are held constant, or whether domain-specific knowledge is primarily driven by acoustic features that are characteristics of each domain. We examined how adults and children detect pitch interval changes in both a spoken and sung context for overtly spoken and sung utterances and for ambiguous utterances that straddle the boundary between speech and song. Using a pitch contour expansion, we expanded the pitch relationships of spoken, sung, and ambiguous utterances, such that the pitch contour was preserved, but pitch interval relationships were broken. Pitch contour is a salient aspect of both speech and song, but pitch interval relationships are uniquely important in music. Thus, if listeners recruited domain-specific knowledge when listening to overtly spoken and sung utterances, listeners should be more sensitive to these pitch interval changes for song than speech. We also examined pitch interval change sensitivity for ambiguous utterances when they were presented in a language vs. a music context. Finally, as domain-dependent recruitment of knowledge can help listeners extract relevant features in each domain, we examined whether domain-specific recruitment of knowledge was correlated with phonological processing. Adults showed greater sensitivity for sung compared to spoken utterances, but their performance was not related to phonological processing. Results for children (4-8 years of age) will be discussed.

Perception of Vowel-Consonant Duration Contrasts in 8-month-old Infants

Christine D. Tsang[1] and Simone Falk[2]
[1] Department of Psychology, Huron University College at Western, London, Ontario, CANADA
[2] Department of German Philology, Ludwig-Maxmillians-University, Munich, GERMANY

In languages such as German and English, differences in meaning can be indicated by distributional properties of between vowel and consonant durations. In words such as “bean”, long vowels are followed by relatively short consonants, while short vowels as in “bin” are followed by relatively long consonant durations. So far, it is unclear if infants are sensitive to such distributional properties of vowel and consonant durations. Moreover, it has been shown with adults and infants that rhythmic structure in speech can aid discrimination of durational properties of segments. In the present study, we investigated the ability of infants to recognize a familiarized vowel-consonant duration contrast in a regular rhythmic speech context compared to a novel contrast. Thirty-two 8-month-old English-learning infants were familiarized for 2 minutes to three spoken versions of a trisyllabic nonword with a strong-weak-weak stress pattern. The nonword comprised two sonorant consonants (/l/,/m/) and two vowels (/i/,/a/). In a first set of stimuli, the vowels were long and the following consonants were short(/’mi:la:mi/, while in a second set of stimuli, the vowels were short and the consonants were long (/’mil:am:i/). Following the familiarization, infants were tested using a head-turn preference procedure on their preference for the familiarized durational pattern or a novel pattern by presenting two novel sets of words with similar stress and segmental structure (i.e., /’la:mi:la/ or /’lam:il:a/). Results indicated that infants showed significantly longer attention to the novel durational pattern. This result indicates that infant listeners 1) are able to abstract durational properties of segmental patterns during a brief familiarization and 2) can recognize the familiar from novel segmental patterns in a rhythmically regular speech context. A follow-up study aims at examining whether irregularity in rhythmic structure impairs recognition of the durational pattern.

A novel auditory classification paradigm for developmental music cognition research

Justin Lane Black[1] and Erin Hannon[1]
[1]Psychology Department, UNLV, Las Vegas, Nevada, USA

Much of behavioral music cognition research relies on the attentional control and self-discipline of the listener, who is asked to make ratings or judgments about a sound, or to classify a sound into one of multiple categories. Although such tasks are easily carried out by adults, they may pose an unnecessary challenge to populations such as young children. While a venerable tradition of "rocket-ship psychophysics" employs cleverly designed games that measure discrimination and detection thresholds among children (Abramov et al, 1984), there is an ongoing need for ecologically valid, engaging tasks to be used in auditory cognition research with children. We adapted a video game paradigm (Lim & Holt, 2011) to examine auditory pattern-learning in adults and children. The fictional premise is that the player (child) must save Earth by capturing or destroying noisy aliens, whose sounds predict their status as friends or enemies before they appear on screen. During a training phase listeners are trained with feedback to classify aliens according to two sound patterns (the "friend" or "enemy" pattern), and during a test phase they must classify aliens in the dark (due to a "power outage"). We have adapted this paradigm to replicate a prior adult pattern-learning experiment in our lab showing that listeners can classify (using similarity ratings) novel syllable or melodic patterns based on prior exposure to other exemplars (Hannon et al., in preparation). Assuming full replication of adult findings is successful, we will extend the task to children. The game is general enough that it could be used to ask many different types of questions involving auditory classification, but by virtue of its ecological validity it promises to yield more robust, reliable data from young children.

Modality and tempo of metronome-timed speech influence the extent of induced fluency effect in stuttering speakers

Lauren Amick[1], Katherine Jones[1], Juli Wade[1,3], Soo-Eun Chang[2], J. Devin McAuley[1,3]
[1]Psychology, Michigan State University, East Lansing, MI, USA
[2]Psychiatry, University of Michigan, Ann Arbor, MI, USA
[3]Neuroscience, Michigan State University, East Lansing, MI, USA

Stuttering is a neurodevelopmental disorder characterized by disfluent speech patterns. Consistent with a framework that proposes that deficiencies in rhythm and timing are one mechanism behind stuttering, researchers have demonstrated that speech fluency in individuals who stutter improves when speech is timed with an external pacing signal, such as a metronome. In this study, we explored the effects of metronome modality and tempo on speech fluency in adults who stutter. In accord with evidence that individuals demonstrate greater beat sensitivity in auditory rhythms than visual rhythms, we hypothesized that an auditory metronome as compared to a visual metronome would induce greater speech fluency in speakers who stutter. In addition, due to a generally greater difficulty synchronizing with slow rhythms compared to fast, we hypothesized that a metronome at a faster tempo would induce greater fluency than a metronome at a slower tempo. Participants produced speech in synchrony with visual or auditory metronomes at fast and slow tempos. We also compared each metronome condition to two baseline measures – natural reading, and self-paced reading. Voice recordings were analyzed by a certified speech-language pathologist and assessed for percent stuttering-like disfluencies. Results revealed an overall effect of metronome-timed speech, such that speech was more fluent when timed with a metronome than during natural reading conditions. Fluency was greater in the auditory condition than the visual condition, but only at the fast tempo. Fluency was no different in self-paced speech conditions than in auditory metronome or slow visual metronome conditions. Less fluent speech in the visual fast condition suggests an interference of the visual metronome with the fluency-inducing effect of paced speech. These results suggest that for adults who stutter, rhythmically pacing speech without an external signal may have the same fluency-inducing effect as metronome-timed speech.

A Study on Analysis of Prosodic and Tonal Characteristics Emotional Vocabularies in Korean Language for Songwriting

Soo Yon Yi[1], Hyun Ju Chong[1]
[1]Department of Music Therapy Graduate School, Ewha Womans University , Seoul, Korea

Song is combined form of words and music, which is efficient medium for one's expression of emotions and thoughts. Songwriting is one of major therapy techniques utilized to facilitate self-expression including diverse emotions. Therefore, selecting and matching vocabularies for the melodic line, or vise versa, is an important concept in songwriting. A few studies on music and language suggest that there is a similarity between a melody of song and a prosodic characteristic of language. However, there are not many studies which examine relationships between the prosodic contours of emotion vocabularies from musical or tonal aspects. These studies can provide crucial information needed for songwriting.

This study was purposed to investigate the characteristic of tonal elements in emotional vocabularies in adults by analyzing the prosodic features of vocabularies used for designated context when speaking: pitch pattern, loudness (dB), and speech rate (syl/sec). For the analyses, words of 3 different emotions from valence theory (happy, sad, fear/anxiety) and one neutral word (be careful) were processed to extract acoustic values using Pratt. The data were transformed into pitch (herz) as the Pratt records them in semitone figures in order to examine the melodic contour.

The results showed that 'ʙul-ŋan-hɛ-sʌ' (fear/anxiety) and 'ʤl-gʌ-ŋwa-ŋjo' (fun) were common pitch patterns among male. The pitch patterns of these syllables are that 'ʙul-ŋan-hɛ-sʌ' is A2-A2-G2-B2 and 'ʤl-gʌ-ŋwa-ŋjo' is G2-G2-G2-E2. Whereas, 'sɯl-pʌ-ŋjo' (sad), 'hɛŋ-bok-hɛ-sʌ' (happy), 'ʤo-sim- hɛ- ŋjo' (be careful) were common pitch patterns among female. The pitch patterns of these syllables are that 'sɯl-pʌ-ŋjo' is A3-B3-A3, 'hɛŋ-bok-hɛ-sʌ' is D4-E4-B3-B3 and 'ʤo-sim- hɛ- ŋjo' is G3-G3-D3-A2. Another results were that the mean loudness of 'sad' is 40 (dB), 'happy' is 44.47 (dB) and 'be careful' is 41.71 (dB). Overall, results suggest that the word 'happy' had the highest pitch pattern and loudness whereas the word 'sad' had lower pitch pattern and loudness. There were differences in the speech rate among the words although not at the significant level. The Results provided basic data on the relationship between emotionality of words and tonal contour. Further, these findings can be incorporated into songwriting.

The association between musical skills and grammar processing in childhood

Swathi Swaminathan[1] and E. Glenn Schellenberg[1]
[1]Department of Psychology, University of Toronto, Mississauga, ON, Canada

In this experiment we investigated whether musical skills are associated with grammar processing in musically trained and untrained 6- to 9-year-old children. The children completed a music aptitude test and a standardized test of receptive grammar. After controlling for demographic factors, working memory and IQ, rhythm-perception skill, but not melody perception, was found to predict receptive grammar. This result suggests an overlap in the temporal domain for language and music. Musically trained children outperformed on the test of melody perception. They did not, however, show a rhythm perception or receptive grammar advantage. This result suggests that music training may not lead to language advantages if it does not promote the development of rhythm-perception skill.

Evaluating the musical protolanguage hypothesis using a serial reproduction task

William Thompson[1,2], Anna Fiveash[1,2] and Weiyi Ma[2]
[1]Department of Psychology, Macquarie University, Sydney, NSW, Australia
[2]ARC Centre of Excellence in Cognition and its Disorders

The origin of music and language is a fundamental question in science. Darwin's musical protolanguage hypothesis holds that music and language evolved from a common precursor that gradually diverged into two distinct systems with distinct communicative functions. In this investigation, a *serial reproduction* task was used to simulate this process of divergence. Fifteen participants were presented with eight vocalized nonsense sounds and were asked to reproduce them with specific communicative intentions. Four vocalizations functioned to communicate an emotional state (happy, sad, surprised, tender) and four vocalisations were referential, communicating a physical entity (tree, hill, stone, grass). Participants were told to reproduce each sound with the defined communicative intention strongly held in mind (the assignment of the eight vocalisations to emotional or referential conditions was counterbalanced across participants). The resultant vocalisations were then recorded and played to a new set of 15 participants, who were given the same instructions as the first set of participants. This procedure was continued across five sets of participants ("generations") to create 15 "chains" of reproductions. Recordings of the *final* 120 vocalisations were presented to two new groups of participants in Australia (N=40) and China (N=40), who rated each in terms of its perceived music- and speech-like qualities. Emotional vocalizations were rated as significantly more music-like than referential vocalisations; and referential vocalisations were rated as significantly more speech-like than emotional vocalizations. Acoustic analysis of vocalizations revealed clear differences in rate, intensity, and pitch variation for emotional and referential vocalizations. This study is the first to experimentally demonstrate how a musical protolanguage may have gradually diverged into speech and music depending on the communicative intention of vocalizations.

Instrumental functionality and its impact on musical perception

Morten Schuldt-Jensen[1]
[1]Department of Conducting, Hochschule für Musik /University of Music, Freiburg, Germany

In order to convey not only the semantic information of a score but also the emotional connotation of the work, every interpretation of a classical composition implies an attempt to identify the artistic and even ethical-religious intentions of the composer. The findings are to be expressed in the performance by means of only a few musical categories - tempo, dynamics, phrasing, articulation, timbre – which on a technical level, however, encompasses a host of optional parameters, among which musicians and conductors must permanently navigate in order to target the emotional impact of their interpretations.

This presentation aims to explain and exemplify these options based on recent findings within the fields of musical perception and semiotics.

As human beings we are especially conditioned to receive emotional messages by means of the human voice, and recent research (Sundberg et al. 2015) has shown, that different vocal registrations consistently elicit certain emotions in the listener, probably based on a bodily/tactile recognition of the spontaneous vocal functionality in a certain emotional state. This inferred maluma-takete phenomenon holds good for – transposes into – the functionality of instrumental music as well.

Drawing on a number of performance practical insights from a conductor's point of view and by analyzing excerpts from highly contrasting recordings of e.g. Bach's St. Matthew Passion and the Requiems of Mozart and Brahms the presentation will show, how a variety of apparently mere functional choices, from string bowings and vibrato frequency to the vocal registration of the soloists, possible disregard of biographical and „less important" score information, as well as ignorance of period performance conventions can dramatically change the intended emotional and theological meaning of a composition – and thus eventually its artistic effect and religious function.

The Conductor as Guide: Gesture and the Perception of Musical Content

Anita B. Kumar,[*1] Steven J. Morrison[*2]

[*]Laboratory for Music Cognition, Culture, and Learning, School of Music, University of Washington, USA

[1]kumarab@uw.edu, [2]sjmorris@uw.edu

I. BACKGROUND

Conducting pedagogy often depicts the conductor as an embodiment of music. The common perception of the visual element of conductor gesture is that it has a direct linear affect on performers' responses and ultimately the perception of the music by the audience. However, our understanding is growing in how the multimodal interaction of visual and auditory information simultaneously impacts the perception of musical performances. Recent research has quantified this phenomenon through examination of the effect of expressive gestures on viewers' evaluative perceptions of conducted ensemble performances. (Morrison et al., 2009; Morrison et al., 2014) Previous research has also documented the way the visual domain interacts with the auditory domain for perception of specific musical elements, including pitch perception (Connell, Cai, & Holler, 2013), consonance/dissonance (Thompson, Graham & Russo, 2009; Thompson & Russo, 2007), personification and character (Graybill, 2011; Gritten, 2011), and how movement is visualized in mental imagery (Eitan & Granot, 2006; Maes et al., 2014). We hypothesize that the effect of visual information from a conductor's movements may be due to conductor gesture delineating and amplifying specific expressive aspects of music performances.

II. PURPOSE

The purpose of the present study is to determine if differing conductor gesture affects observers' focus of attention to contrasting aspects of ensemble performances.

III. METHOD

Two excerpts were arranged for a small conducted ensemble of seven players. Each excerpt featured two-part counterpoint of contrasting elements: (1) an ostinato paired with a lyric melody and (2) long chord tones paired with rhythmic interjections. Because of listeners' tendency to focus on higher pitched lines, each excerpt was audio recorded in two forms in which the paired elements were alternated between upper and lower voices. Audio was recorded separately to ensure both lines are presented at equivalent amplitude.

Audio recordings were paired with video of two different conductors. Each conductor used gesture appropriate to one or the other musical element for a total of sixteen videos, ordered in two equivalent test forms. In each form participants heard the same voiced excerpt twice, once for each gesture condition (congruent with the upper or the lower line) for a total of eight items. Presentation order was such that no conductor appeared consecutively. Participants responded to a questionnaire in which they are asked to evaluate each excerpt along 10-point differential scales anchored by descriptive terms. These descriptions focused on elements of articulation, style, rhythm regularity, and phrase duration. Participants also indicated whether a lower timbre (Tuba) or a higher timbre (Flute) were dominant in the excerpt.

IV. RESULTS

Audio recordings were paired with video of two different conductors. Each conductor used gesture appropriate to one or the other musical element for a total of sixteen videos, ordered in two equivalent test forms. In each form participants heard the same voiced excerpt twice, once for each gesture condition (congruent with the upper or the lower line) for a total of eight items. Presentation order was such that no conductor appeared consecutively. Participants responded to a questionnaire in which they are asked to evaluate each excerpt along 10-point differential scales anchored by descriptive terms. These descriptions focused on elements of articulation, style, rhythm regularity, and phrase duration. Participants also indicated whether a lower timbre (Tuba) or a higher timbre (Flute) were dominant in the excerpt.

Overall, we observed differences in evaluations of Excerpt 1 (lyric melody and ostinato) according to expectation. When conductor gesture was congruent with the ostinato, participants rated the excerpt as having more *disconnected* Articulations, *irregular* Rhythm, *angular* Style, and *short* Phrases. When gesture was congruent with the melody, ratings indicated more *connected* Articulation, *regular* Rhythm, *flowing* Style, and *long* Phrasing. This result was evident regardless of arrangement (with the melody performed by either the upper voices or lower voices, labeled "MO" and "OM", respectively, in Fig. 3.) Ratings of Excerpt 2 (chords with interjections, labeled "CI" and "IC") showed a less consistent relationship with gesture, with the arrangement featuring upper voice chords and lower voice interjections ("CI") evoking evaluations contrary to expectation.

We used a mixed linear model to determine the degree to which gesture, conductor, and excerpt arrangement, as well as interactions between gesture/conductor and gesture/excerpt arrangement, were factors in ratings for each of the four dichotomous rating scales. The model confirmed a significant difference between excerpts and within excerpts between voicing arrangements. Although there was a significant difference between conductors on the domains of Rhythm and Style, this difference did not interact with gesture.

There were moderate positive correlations between Articulation and Style scores (r = 0.60, p = 0.01), and Style and Phrasing scores (r = 0.51, p = 0.01). There were no significant differences for gesture in Excerpt 2 (chords/interjections) in any of the four domains. This result is not surprising due to the contrasting response patterns between the two voicings of this excerpt (Figure 3). No significant effects were observed for the domains of Rhythm (p = 0.546 for both excerpts) and Phrasing (p = 0.053 for Excerpt 1, and p = 0.15 for Excerpt 2) for either excerpt, though responses in these domains generally resembled those observed for Articulation and Style.

V. CONCLUSIONS

The purpose of this study was to determine if observers' perceptions of musical content were related to conductor gesture. Previous research has documented the powerful role of visual information in the evaluation of music performances (Tsay, 2013). Most of the findings related to this topic have examined the broad construct of expressive movement and have compared performances either with or without accompanying expressive gesture (e.g., Morrison et al., 2009), with varied levels of expressive gesture (e.g., Vines, Krumhansl, Wanderley, Dalca & Levitin, 2011), or with varied levels of performance quality (e.g., Silvey, 2011). Here, rather than examine the presence, absence or relative magnitude of expressive gesture, we considered whether the content of musical gesture was related to the way in which one understood a music performance. We predicted that a conductor's gestural focus on particular aspects of a piece of music would affect the way an observer described the performance.

The present data suggest that conductor gesture does appear to have an effect on observer perception. Consistent with previous research, the visual aspect of conducting plays a consequential role in the perception of a conducted ensemble's performance (Madsen, Geringer & Wagner, 2007; Madsen, Geringer & Madsen, 2009; Price & Mann, 2011). The relationship between conductor gesture and observer description was particularly evident in the domains of articulation and style, aspects of music performance that have been reported elsewhere to yield notably different responses from listeners depending on the movement accompanying them (Morrison et. al, 2014). Rather than being consistent in its impact, however, this effect appears to be mediated by preconceptions of the importance of melodic over rhythmic material, of certain timbres over others, and of the durations between onsets of new active material. This suggests that visual information provided by conductor gesture is among the resources viewers draw upon to make sense of what they perceive as part of the performance. Music performances are therefore a multimodal experience, with visual and auditory information comingled in how music is experienced.

REFERENCES

Connell, L., Cai, Z. G., & Holler, J. (2013). Do you see what I'm singing? Visuospatial movement biases pitch perception. *Brain and Cognition*, *81*, 124–130. http://doi.org/10.1016/j.bandc.2012.09.005

Eitan, Z., & Granot, R. Y. (2006). How Music Moves. *Music Perception*, *23*, 221–248. http://doi.org/10.1525/mp.2006.23.3.221

Gritten, A. Distraction in polyphonic gesture. (2011). In Gritten, A., & King, E. (Eds.) *New perspectives on music and gesture*. Farnham; Burlington, VT: Ashgate Pub.

Graybill, R. (2011). Whose gestures? Chamber music and the construction of permanent agents. In Gritten, A., & King, E. (Eds.) *New perspectives on music and gesture*. Farnham; Burlington, VT: Ashgate Pub.

Madsen, C. K., Geringer, J. M., & Madsen, K. (2009). Adolescent musicians' perceptions of conductors within musical context. *Journal of Research in Music Education*, *57*(1), 16–25. doi:10.1177/0022429409333352

Madsen, C. K., Geringer, J. M., & Wagner, M. J. (2007). Context specificity in music perception of musicians. *Psychology of Music*, *35*(3), 441–451. doi:10.1177/0305735607072655

Maes, P., Leman, M., Palmer, C. & Wanderley, M. M. (2014) Action-based effects on music perception. *Frontiers in Psychology, 4: 1008. doi: 10.3389/fpsyg.2013.01008*

Morrison, S. J., Price, H. E., Geiger, C. G., & Cornacchio, R. A. (2009). The effect of conductor expressivity on ensemble performance evaluation. *Journal of Research in Music Education*, *57*(1), 37–49. doi:10.1177/0022429409332679

Morrison, S. J., Price, H. E., Smedley, E. M., & Meals, C. D. (2014). Conductor gestures influence evaluations of ensemble performance. *Frontiers in Psychology*, *5*, 806. http://doi.org/10.3389/fpsyg.2014.00806

Price, H. E., & Mann, A. (2011). The effect of conductors on ensemble evaluations. *Bulletin of the Council for Research in Music Education, 189,* 57-72.

Silvey, B. A. (2011). The effect of ensemble performance quality on the evaluation of conducting expressivity. *Journal of Research in Music Education*, *59*(2), 162–173. doi:10.1177/0022429411406173

Thompson, W. F., Graham, P., & Russo, F. A. (2005). Seeing music performance: Visual influences on perception and experience. *Semiotica*, *2005*, 203–227.

Thompson, W. F., Russo, F. A. (2007). Facing the music. *Psychological Science, 18,* 756-757.

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., Dalca, I. M., & Levitin, D. J. (2011). Music to my eyes: Cross-modal interactions in the perception of emotions in musical performance. *Cognition*, *118*, 157–170. http://doi.org/10.1016/j.cognition.2010.11.010

How Music Moves Us: Entraining to Musicians' Movements

Roger Chaffin[1] & Alexander P. Demos[2]
[1]Department Psychology, University of Connecticut, Storrs, CT, USA
[2]Department Psychology, University of Illinois at Chicago, IL, USA

Why do listeners move to music? To find out, we measured the postural sway of two trombonists as they each recorded multiple performances of two solo pieces in each of three different expressive styles (normal, expressive, non-expressive). We then measured the postural sway of 29 non-trombonist listeners as they "air conducted" the performances (Experiment 1), and of the two trombonists as they played along with the same recorded performances (Experiment 2). In both experiments, the velocity of listeners' sway was more similar to that of the performer than expected by chance. In other words, listeners spontaneously *entrained* to the performer's movements, spontaneously changing their direction of sway at the same time as the performer simply from hearing the recorded sound of the music. Listeners appeared to be responding to information about the performer's movements conveyed by the musical sound. First, entrainment was only partly due to performer and listeners both swaying to the musical pulse in the same way; listeners also entrained to performers' unique, non-metrical, movements. Second, listeners entrained more strongly to back-and-forth sway (more closely linked to the sound-producing movements of the trombone slide) than to side-to-side sway (more ancillary). Both effects suggest that listeners heard the performer's movements in the musical sound and spontaneously entrained to them. People spontaneously entrain to the movements of other people in many situations. In present study, entrainment appeared to be an integral part of listeners' affective response to the music. Listeners rated performances as more expressive when they entrained more. They also rated expressive higher than non-expressive performances, but they did not entrain more strongly to more expressive performances. Thus, entrainment was determined by the listener, not by the expressive style of the performer. Entrainment appears to enhance listeners' experience of musical feelings expressed by the musician.

The effect of attention on adaptive and anticipatory mechanisms involved in sensorimotor synchronization

Bronson Harry[1], Carla Pinto[1], and Peter Keller[1]
[1]The MARCS Institute, Western Sydney University, Sydney, New South Wales, Australia

Human interaction often involves the rhythmic coordination of actions across multiple individuals, such as in ensemble musical performance. Traditional accounts of sensorimotor synchronization (SMS) posit that precise synchronization is the product of adaptive mechanisms that correct for asynchronies between performers in a reactive manner. However, such models cannot account for how synchrony is maintained during dynamic changes in tempo. The adaptation and anticipation model (ADAM) proposes that performers also make use of predictive mechanisms to estimate the timing of upcoming actions via internal simulations of a partners' behavior. According to this account, predictive mechanisms rely on attention whereas basic adaptive mechanisms are automatic. To investigate the effect of attention on adaptive and anticipatory mechanisms, we fitted a computational implementation of ADAM to data collected from two SMS tasks. Estimates of adaptation were derived from a SMS task involving an adaptive virtual partner, wherein participants drummed in synchrony with an adaptive metronome that implemented varying levels of phase correction. Estimates of anticipation were derived from a task in which participants drummed in synchrony with sequences with constantly varying tempo. To assess the role of attention, participants also did these tasks while engaged in a concurrent visual one-back task. As expected, synchronization accuracy and precision were not modulated by the secondary task for the virtual partner task, but were significantly impaired under dual task conditions for the tempo-changing task. Moreover, model estimates of adaptive mechanisms were not altered by the secondary task, whereas model estimates of predictive mechanisms indicated participants utilized reduced predictive control while engaged in the secondary task. These results suggest that rhythmic interpersonal synchronization relies on both automatic adaptive and controlled predictive mechanisms.

Gesture, Embodied Cognition, and Emotion: Comparing Hindustani Vocal *Gharanas* in Performance

Hans Utter
[1]Ohio Arts Council Artist in Residence, USA; email: hansutter@hotmail.com

Introduction

Bilateral gesture in North Indian classical vocal music (khyal) has a complex relationship with music structure and aesthetic intentions, and appears to augment emotional expression and serve important pedagogical functions. Gesture has an important role in Indian vocal training, especially in relation to processes of attention, cognition, and memory. In this paper, I examine the multimodal interactions between voice production, emotional expression, song text, and gesture through an audio/visual analysis of performers in three different vocal traditions. I will examine distinctions between different vocal *gharanas* (schools) through both gesture and related stylistic/aesthetic elements. The use the spatial visualization of sound based on particular vowels in the *dhrupad* vocal genre will also be discussed as a possible influence on movement.

Method

The methods employed consist of collecting ethnographic data through interviews and audio/video recordings, participation in learning situations, and examining Indian theoretical treatises on vocal music. The data is correlated with performance practices of the *gharanas* to determine the relationships between aesthetics, vocal culture, and musical structure.

Results

While the use of gesture is not formally taught and can be idiosyncratic, performers from specific vocal *gharanas* exhibit some common practices. However, between *gharanas* gestures tend to emphasize the dominant aesthetic features particular to that *gharana,* and correlate with distinct musical elements.

Discussion

For North Indian classical vocal music, the system of oral transmission and individualized training in a particular *gharana* is evident in gesture, which developed through the process of enculturation. My evidence shows that gesture is a component of multi-modal processes of learning and expression, which supports the hypothesis of the embodied nature emotional-cognitive processes.

The effects of ancillary gestures on interpersonal coordination in ensembles

Petri Toiviainen[1] and Peter Keller[2]
[1]Department of Music, University of Jyväskylä, Finland
[2]MARCS Institute, Western Sydney University, Australia

Ensemble performance entails the coordination of sound-producing instrumental movements, while relatively unconstrained ancillary gestures communicate expressive intentions and regulate co-performer interaction via the visual modality. Previous work has shown that the interpersonal coordination of instrumental movements and ancillary gestures is correlated, but little is known about this relationship and whether it is influenced by musical roles (melody vs. accompaniment), which may affect leader-follower relations. We sought to identify features of ancillary gestures that influence the interpersonal coordination of instrumental movements, and to determine whether the relative importance of these features varies with musical role (primo vs. secondo) in duos. Data from 16 pairs of pianists were analyzed. From a collection of 8 piano duets, each pair performed 2 duets, 6 times each, without visual contact on electric keyboards, while their upper body movements were recorded with a motion capture system. Subsequently, for each performance, interpersonal keystroke synchrony was quantified from the MIDI data and the motion capture data were decomposed into eigenmovements by means of a group Principal Components Analysis. Finally, multiple regression analyses were applied to predict keystroke synchrony from the interpersonal mutual information in the eigenmovements and the energies thereof. The degree of keystroke synchrony could be significantly predicted both by the mutual information and by the energy. In particular, high mutual information between the players' anteroposterior body sway and head nodding predicted high synchrony. Moreover, high energy in the secondo player's anteroposterior movement was associated with high keystroke synchrony. The results suggest that ancillary anteroposterior body sway facilitates the interpersonal coordination of instrumental movements, hence sounds, in ensembles. Performers playing a follower role contribute more, presumably by adapting to the leader's timing.

Connecting conductor gesture to compositional features and conductors' expressive intentions: an exploratory kinematic study

Yu-Fen Huang[1], Nikki Moran[2], Raymond MacDonald[2], Simon Coleman[3]
[1]Institute for Music in Human and Social Development, Reid School of Music and Institute for Sport, Physical Education and Health Sciences, Moray House School of Education, Edinburgh University, Edinburgh, UK
[2]Institute for Music in Human and Social Development, Reid School of Music, Edinburgh University, Edinburgh, UK
[3]Institute for Sport, Physical Education and Health Sciences, Moray House School of Education, Edinburgh University, UK

Orchestral conductors use body movement to communicate their interpretation of musical compositions, relaying to the orchestra their intentions regarding the way that both intrinsic and expressive structural features (melodic line, harmonic progression, metre, dynamic range, etc.) are delivered in performance. The communicative nuance of conductors' gestures is much richer than is suggested by general rules addressed in conducting manuals; meanwhile, accounts which address detailed interpretation tend not to address the pragmatic detail of conducting movement. This study explores how kinematic features of six conductors connect to compositional features, focusing on the relationship of certain movements to the unfolding temporal structure of the performance; and also to features selected by each individual conductor according to their own interpretation and communicative intentions. Six conductors worked with a small string ensemble, rehearsing and recording excerpts from three pieces of music by Mozart, Dvořák and Bartók. Their conducting movements were recorded by an optical motion capture system Qualisys. Kinematic features including movement distance, speed, acceleration and jerk were extracted using Visual 3D. Correlational analysis revealed distinctive kinematic features of conducting movement within each item of repertoire. The movements showed higher variability at the musical events highlighted by the individual conductors, with conductors demonstrating multiple strategies to modulate those musical features they intended to highlight. Temporal analysis focused on specific time points where distinctive movement features are identified, and – based on qualitative analysis of interview data with the conductor participants – explore the relationship between these movement changes and conductors' stated intentions regarding their interpretation of the musical compositions.

The Sounds of Movements: Self-Other Perception of Musical Gestures in Multimodal Conditions

Jesper Hohagen,[*1] Clemens Wöllner,[*2]

[*]*Institute of Systematic Musicology, University of Hamburg, Germany*
[1]jesper.hohagen@uni-hamburg.de, [2]clemens.woellner@uni-hamburg.de

STRUCTURED ABSTRACT

A. Background

Truslit (1938) developed a theory on the gestural quality of musical interpretations. Since Bruno Repp's (1993) synopsis of Truslit's work, researchers have increasingly developed methods and paradigms to study listeners' responses to musical gestures. Truslit originally presented movement graphs for selected musical pieces, based on dynamic and agogic information, assuming that these motion trajectories are valid intersubjectively. Further empirical examinations are required to test this theory of a prototypicality of music-inherent movements on a descriptive and comparative level.

Self-other judgment paradigms of visual point-light movements (Sevdalis & Keller, 2010; Wöllner, 2012) allow elucidating action-perception coupling processes (Prinz, 1997) underlying musical performance movements as described by Truslit. In addition, movement sonification appears to be a promising methodological approach in this regard. So far, movement sonification has mainly been applied in artistic performances (Renault et al., 2014) or sport and rehabilitation science (Schaffert, 2011; Effenberg et al., 2011). To our knowledge, our study is the first to systematically investigate the motion of Truslit's gestures with movement sonification and a self-other perception paradigm.

B. Aims

In a first study, we investigated Truslit's hypothesis of prototypical musical gestures by comparing free movements to Truslit's original sound examples with movements following a visual presentation and detailed verbal instructions. The second study tested the effects of watching point-light displays and listening to the sonification of movements with a multimodal self-other judgment paradigm.

C. Method

In Study 1, a total of 26 right-handed participants (age: M=27.35, SD=4.06; 30,8% female, 50% musicians) were tracked with a motion capture system while executing arm movements along with Truslit's (1938) original musical examples.

The second study part consisted of a self-other perception judgment paradigm, presenting sequences to the same participants (matched with those of two other participants, unbeknown to them) under four different conditions: 1) visual 2D point-light display: "v", 2) sonification: "a", 3) 2D drawing map without movements or sounds (still image: "si"), 4) combination of conditions 1 and 2: "av".

C. Results

In order to assess differences between free and post-instruction movements within Study 1 as well as between the two musical excerpts, we analyzed the averaged global measures of their index finger movement lines in terms of movement velocity, acceleration, jerk and cumulative distance. A 2x2 repeated-measures ANOVA indicated significant differences in the velocity values between free movements and movements following an instruction and visual presentation ($F\{1, 25\}=5.40$, $p=.029$, $\eta^2=.177$), indicating that participants moved more quickly in the free condition compared to the post-instruction condition.

Analyses of the self-other recognition task in Study 2 addressed judgment sensitivity by calculating d-prime (d') scores for individual participants. One-sample t-tests revealed that self-recognition was successful in three conditions: v ($t\{22\}=2.21$, $p<.05$); si ($t\{22\}=2.45$, $p<.05$) and av ($t\{22\}=2.46$, $p<.05$). Furthermore, musicians scored significantly higher in the free visual ($t\{21\}=2.29$, $p<.05$) and free audiovisual conditions ($t\{21\}=2.31$, $p<.05$) compared with non-musicians.

E. Conclusions

While in Study 1 there were large inter-individual differences in the movement trajectories between participants, analyses revealed a high consistency in the repeated-measures condition, so that individuals performed comparable movements across trials. However, significant differences between performance conditions (free – after instruction) indicate small effects of moving intuitively on movement velocity.

Results of the self-other recognition (Study 2) suggest that the visual information of the movement trajectories predominantly carried information for successful recognition, comparable to previous research (Sevdalis & Keller, 2010; Wöllner, 2012). Musicians recognized their movements correctly more often compared to non-musicians, possibly showing advanced musical perception processes due to their expertise in moving while making music. D-prime scores for the audio condition showed that participants could not recognize auditory displays of their movements above chance. In a further study, we will create and evaluate different auditory display methods, aiming at studying various sonification dimensions of movement data.

REFERENCES

Effenberg, A.O., Fehse, U., Weber, A. (2011). Movement Sonification: Audiovisual benefits on motor learning, In B.G. Bardy, J. Lagarde, D. Mottet (eds.) *BIO Web of Conferences. The International Conference SKILLS, pp 1-5.*

Prinz, W. (1997). Perception and Action planning. *European Journal of Cognitive Psychology, 9*, 129-154.

Renault, A., Charbonnier, C., Chagué, S. (2014). 3dinmotion – a Mocap Based Interface for Real Time Visualisation and Sonification of Multi-User Interactions, In *NIME International Conferences on New Interfaces for Musical Expression, pp. 495-496.*

Repp, B. (1993). Music as Motion: A Synopsis of Alexander Truslit´s (1938) Gestaltung und Bewegung in der Musik. *Psychology of Music, 21*, 48-72.

Schaffert, N. (2013). *Sonifikation des Bootsbeschleunigungs-Zeit-Verlaufs als akustisches Feedback im Rennrudern*, Berlin: Logos.

Sevdalis, V., Keller, P.E. (2010). Cues for self-recognition in point-light displays of actions performed in synchrony with music. *Consciousness and Cognition, 19*, 617-626.
Truslit, A. (1938). *Gestaltung und Bewegung in der Musik: Ein tönendes Buch vom musikalischen Vortrag und seinem bewegungserlebten Gestalten und Hören.* Berlin-Lichtenfelde: Vieweg.
Wöllner, C. (2012). Self-recognition of highly skilled actions: A study of orchestral conductors. *Consciousness and Cognition, 21*, 1311-1321.

The loss and regain of coordinated behavior in musical duos

Andrea Schiavio[1,2], Ashley Walton[3], Matthew Rodger[4] and Johanna Devaney[1]
[1] School of Music, The Ohio State University, Columbus, OH, USA
[2] Department of Music, The University of Sheffield, Sheffield, UK
[3] Department of Psychology, University of Cincinnati, Cincinnati, OH, USA
[4] School of Psychology, Queen's University Belfast, Belfast, UK

Traditionally, participatory musical behavior has been explored through what seems to be an individualist perspective, which considers the single agent as the main explanatory unit of the interaction. This view, however, may downplay the role of real-time interactions, reducing joint action to a stimulus-response schema. Our aim, then, is to go beyond such spectatorial stance, investigating how musicians maintain structural unity through a mutually influencing network of dynamical processes. For example, is visual information needed in order to achieve coordination among musicians? What kind of coupling is necessary to keep playing together when the auditory signal is perturbed externally? In order to answer such questions, we will recruit 6 pairs of string players divided into 2 groups (by level of expertise) of 3 pairs each. The participants, wearing noise-cancelling headphones, will be firstly asked to play 2 pieces individually (taken from Bartok's 44 Duos). After this, they will perform together in 3 main conditions, designed to test how auditory perturbations may shape coordination when partners are differently 'present' to each other. Conditions involve subjects being (A) naturally placed one in front of the other, (B) able to see each other only partially, and (C) unable to see each other. For all conditions 2 different playing modalities will be investigated: (i) without disturbance, and (ii) with brief, unpredictable, audio disturbance (white noise) in both performers. To investigate interactive behavior, audio and motion-capture data will be extracted, measured, and compared within and between groups during individual and joint performance. Features to be examined include zero-crossing rate, amplitude and duration variables for each movement, and onset synchronicity in the audio signal. We predict significant differences in interacting behavior related to musical expertise, condition, and manipulation.

Impact of event density and tempo on synchronization ability in music-induced movement

Birgitta Burger[1], Justin London[2], Marc R. Thompson[1], Petri Toiviainen[1]
[1]Finnish Centre for Interdisciplinary Music Research, Department of Music, University of Jyväskylä, Finland
[2]Department of Music, Carleton College, Northfield, MN, USA

While most rhythm synchronization studies have focused on finger tapping and stimulus tempo, full-body movements to music can reveal more complex relationships between rhythmic structure and sensorimotor behavior. This study expands on the role of low-frequency spectral flux on synchronization in movement responses to music. The amount of spectral flux in low frequency bands, manifesting in beat-related event density, has previously been found to influence music-induced movement; here, its influence on full-body movement synchronization is examined. Thirty participants were recorded using optical motion capture while moving to excerpts from six Motown songs (105, 115, 130 bpm), three of which containing low and three high spectral flux. Each excerpt was presented twice with a tempo difference of ±5% bpm. Participants were asked to dance freely to the stimuli. Synchronization (phase locking) indexes were determined for the hip, head, hands, and feet, concentrating on vertical acceleration relative to beat-level periods of the music, and sideways acceleration relative to bar-level periods. On the beat level, significant differences were found between high and low flux excerpts, with tighter synchronization of hip and feet to the high flux stimuli. Synchronization was better with the slowed-down stimuli, and decreased in particular for the low flux sped-up excerpts. On the bar level, significant differences between high and low flux excerpts were also found, but with tighter synchronization of head and hands to the low flux stimuli. Moreover, synchronization ability decreased for the sped-up excerpts. These results suggest a connection between event density/spectral flux and phase-locking abilities to music, with an interaction between metrical levels present in the music, spectral flux, and the part(s) of the body being synchronized.

Embodied decision-making in improvisation: the influence of perceptual-motor priming

Eric Clarke[1] and Adam Linson[2]
[1]Faculty of Music, University of Oxford, Oxford, UK
[2]Institute for Advanced Studies in the Humanities, The University of Edinburgh, UK

In this paper we argue for the importance of perception-action coupling in improvisation, and for the more general role of embodiment. We report a study using a novel method to investigate the impact of motor priming on musicians' improvised continuations of ambiguous starting material.

We report a study using a novel priming method to investigate the impact of motor 'shaping' on musicians' improvised continuations from an ambiguous starting point. Trained musicians (N=34) played three games with a Disklavier piano. In Game 1 participants were presented with a short extract of music, automatically played by the piano, which could be interpreted as either a sequence of 2nds or 7ths, and were instructed to improvise a continuation, based on these starting events. In Game 2 they were told to immediately repeat whatever the piano played, as quickly as possible. There were two versions of Game 2 to which participants were randomly assigned in equal numbers: dyads over the whole range of the keyboard that were all 2nds; and dyads that were all 7ths.On completion of Game 2, participants played Game 3, which was identical to Game 1. The dependent variable was the value of the change in the size of the intervals in participants' spontaneous improvised continuations in Game 3 by comparison with Game 1.

The results show an effect of the two different versions of Game 2, with participants exposed to the 7ths version of Game 2 showing an increase in the mean size of the intervals in Game 3 by comparison to Game 1; and participants who were exposed to the 2nds version of Game 2 showing a decrease in the size of the intervals in Game 3 by comparison to Game 1. These results indicate the significant effect of motor priming on improvised musical performance, on the basis of which we argue for the impact of embodied factors on the structure and shaping of musical improvisation.

Embodiment in Electronic Dance Music

Marc Thompson[1] & Hauke Egermann[2]
[1]Finnish Centre for Interdisciplinary Music Research, Department of Music, University of Jyväskylä, Finland
[2]York Music Psychology Group, Department of Music, University of York, York, UK

Electronic Dance Music (EDM) is a predominant cultural phenomenon of every day lives of many. It is listened to in private settings, it contributed to the establishment of club dance culture, and it is also being performed live on big stages that attracts large audiences. However, up until now, it has rarely in the focus of research within the field of music psychology. Given the fact that it is strongly rooted in an embodied engagement with music, its research can be nicely linked to recent theories of embodied music cognition. This symposium will try to contribute to this by providing several very presentations that focus on several stages of the musical communication process within electronic dance music. The first two talks will focus on the production of electronic dance music: One will present the results of a qualitative interview study on live performance of electronic music. Another will present a statistical analysis of historical trends in the production of electronic dance music. Subsequently, the following presentation will focus on audiences moving to this music. This studies will employ an embodied paradigm to research the shape dance movements in club cultural settings. We believe that this symposium will contribute to establish a research focus on this often ignored part of everyday life musical culture.

Embodiment in Electronic Dance Music: The Performer's Perspective

Alexander Förstel[1] & Hauke Egermann[2]
[1]Audio Communication Department, Technische Universität Berlin, Berlin, Germany
[2]York Music Psychology Group, Department of Music, University of York, York, UK

Electronic dance music (EDM) is among the most popular musical genres of today. However, in stark contrast to practices in other musical styles, electronic music is hardly ever played fully “live” in a conventional sense. The performers interact with a variety of electronic devices (e.g. turntables, synthesizers, gestural controllers), often with the use of pre-recorded material, to transform their respective musical intentions into sound. In order to investigate the many possible approaches to electronic “live” performance and the underlying concepts of musical embodiment, 13 qualitative interviews with electronic live performers – ranging from experimental to relatively popular – were conducted. The questions were conceived in the framework of “embodied music cognition”, with a particular focus on the use of the performer's body during the interaction with the equipment, the visibility and readability of the performer's movements from the audience's perspective, and the role of the musical equipment as a mediator between performer and audience. Results initially show a great heterogeneity in the various approaches in the sample group. On closer inspection, a continuum becomes apparent – ranging from a very controlled (i.e. focus on reproduction quality), less embodied and less visible performance style to a very physical, extroverted way of performing with room for unpredictable events and error. Surprisingly, the former “controlled” style is found more often in performers with a connection to club culture, whereas the latter “extroverted” style is found more often in avant-garde contexts with a seated audience. This seemingly contradicts the idea of a shared bio-mechanic energy level between the performer and the audience, as can be found in other popular musical genres. These findings, as well as the general obstacles of conducting research on electronic music, are discussed.

A long term study of tempo in contemporary dance music (1999–)

Dirk Moelants[1,2]
[1]Conservatory, School of Arts, University College Ghent, Ghent, Belgium
[2]IPEM-Department of Musicology, Ghent University, Ghent, Belgium

Several studies have shown a preference for tempi around 120 bpm. If this corresponds to our natural way of moving, it should also be reflected in the tempi found in dance music. To study this, we analyzed the tempi of pieces in the Scandinavian Dance Charts, which give a top 40 of the most popular pieces for each week, including tempo. Analyzing data over a period of 17 years (with 33141 records in total) gives interesting insights in how tempo evolves and is influenced by trends. It shows periods with a prevalence of slower or faster tempi, but most remarkable is the predominance of one single tempo. Over the whole period about 28% of the records have a tempo of 128 bpm, with 62% of the tempi between 125 and 130. This narrow tempo range is always present, independent of certain trends in musical style. Therefore it is a strong indication that that this range corresponds to our natural dance movement and can be related to optimal walking tempo. These results show again a preferred tempo around 120 bpm, but not just 'around'. It seems that 128 bpm has a special function and that a narrow band around it accounts for a majority of our common dance moves. Future research can now consciously look if 128 bpm also has a special role in other rhythmic phenomena and if this might be a more exact representation of human natural resonance frequency.

Embodiment in Electronic Dance Music: How EDM shapes body movement

Birgitta Burger[1]
[1]Finnish Centre for Interdisciplinary Music Research, Department of Music, University of Jyväskylä, Finland

Electronic dance music (EDM) could be classified as music created to move to. While research has revealed relationships between movement features and, e.g., musical, emotional, or personality characteristics, systematic investigations of genre differences and specifically of EDM are rather rare. This contribution aims at offering insights into the embodiment of EDM from three different angles: firstly from a genre-comparison perspective, secondly by comparing different EDM excerpts with each other, and thirdly by investigating embodiments in one specific EDM excerpt. Sixty participants moved freely to 16 excerpts of four different genres (EDM, Latin, Funk, Jazz – four excerpts/genre) while being recorded with optical motion capture. Subsequently, a set of movement features was extracted from the mocap data, including acceleration of different body parts, rotation, complexity, or smoothness. In order to compare embodiments of EDM to the other genres, repeated measures ANOVA analyses were conducted revealing that participants moved with significantly higher acceleration of the center of the body, head, and hands, overall higher smoothness, and a more upright position to the EDM excerpts than to the other excerpts. Between the EDM excerpts, several spectro-timbre related musical features (brightness, noisiness, spectral centroid, spectral energy (mfcc1), attack length) showed high correlations with vertical foot speed, acceleration of head and hands, and upper body complexity. Detailed analysis was conducted on a section containing four bars chorus, four bars break, and four bars chorus of one of the EDM excerpt. Participants' movements differed in several features distinguishing the break from the surrounding parts, in particular showing lower amounts of foot and hand acceleration, phase coherence, rotation and hand displacement in the break. These analyses proposed different ways of studying EDM and could indicate distinctive characteristics of EDM embodiment.

Animated performance: 'Better' music means larger movements

Caitlyn Trevor,[*1] David Huron[*2]

[*]Music, Ohio State University, Columbus, U.S.A.
[1]trevor.5@osu.edu, [2]huron.1@osu.edu

ABSTRACT

Apart from formal dress and composure, musicians give relatively little consideration to visual elements of performance under the assumption that the aural experience far outweighs the visual. However, recent studies suggest that visual information may play a stronger role in aesthetic judgments of musical performances than previously thought. Upon reviewing recent research, we hypothesized that participants viewing musical performances would express a preference for larger performer motions. To test this hypothesis, we used motion capture technology to record four live solo performances with cello, violin, flute, and clarinet. From each original recording, three stick-figure animations were created: one with augmented performance motion, one with the original motion, and one with diminished motion. The three animations were combined into single dynamic videos that allowed participants to adjust the range of motion in the animation via a slider. The slider modified the range of movement continuously from diminished through original to augmented motion. Participants were instructed to adjust the overall amount of performance motion to create the best musical performance. Consistent with our hypothesis, participants elected to significantly augment the motions of the performers. Future studies will investigate more specific elements of expressive movement in performance aesthetics.

I. INTRODUCTION

Musicians are generally aware of potential biases that arise from visual aspects of performance—such as dress, attractiveness, and composure. Some researchers have empirically demonstrated the effects of such biases. For example, Wapnick *et al*. (1997, 1998), found that the attractiveness of a performer can influence judgments of performance quality. However, visual elements are widely considered secondary considerations compared with the purely aural aspects of the performance. In other words, it is often believed that the aural perception of the performance will not be greatly affected by visual aspects of the performance.

Apart from audience attitudes arising from visual cues, some research implies that visual information can potentially alter what a listener actually hears. In a study of multimodal perception, Schutz and Lipscomb (2007) recorded video and audio clips of contrasting articulations on marimba: a short, sharp stroke, and a nominally longer, legato stroke. When the videos of the strokes were switched with each other's audio recording, participants reported hearing longer or shorter tones based solely on the accompanying video clip. In short, the visual action of the mallet stroke altered what listeners heard.

Given that visual information can have a greater impact on auditory perception than is commonly believed, we investigate other visual aspects of performance that might affect listeners' aesthetic judgments.

One consistently present visual aspect of performance is expressive performer movement. Music theorists, music educators, and musicologists have long discussed gesture and motion in music (Berry, 2009; Fisher & Lochhead, 2002; Hatten, 2004; Larson, 2012; McCreless, 2006; Montague, 2012). Morrison *et al*. (2009), found that wind ensemble performances are judged to be more expressive when the ensemble's conductor is more expressive. Similarly, Juchniewicz (2008) compared ratings of a piano performance with amounts of expressive motion and showed that performance elements such as dynamics, phrasing, and overall musical performance are rated higher when performer motion increases. These studies suggest that the amount of movement a performer employs may positively influence judgments of the performance quality.

Intrigued by these findings, we aimed to test whether these results could be replicated in a method of adjustment paradigm. For this study, we hypothesized the following: *(H1): When instructed to adjust the overall motion of an animation of a musical performance to improve the performance, participants will augment the overall amount of performer motion.*

To create a method of adjustment design, we elected to use motion capture technology and animation. This design will be outlined in the methodology section. First, the process of motion capturing to create the main study materials will be outlined.

II. CREATING MATERIALS FOR THE MAIN STUDY: MOTION CAPTURES

A. Participants

Four musicians from Ohio State University, two DMA performance candidates (clarinet and violin) and two faculty members (cello and flute), volunteered to be recorded for the materials for the main study. In order to investigate a wide range of motion capabilities afforded by different instruments, we initially planned to capture one performance from each of the four main instrument families: a string instrument, a woodwind instrument, a brass instrument, and a percussion instrument. However, because the study was conducted over the summer months, recruitment of high-level performers proved difficult. Consequently, our volunteer performers comprise two woodwind players and two string players. Because the cellist and clarinetist would play seated and the violinist and flutist would play standing, a variety of performance motion would nevertheless be presented.

For the capture session, the musicians were instructed to bring two, already-learned musical passages that were each about one minute in length. In considering performer motion, it is appropriate to propose that different styles of music would affect the amount of performer motion both employed by the musicians and preferred by the participants. Because of the difficulty of a fast, technical passage, it is possible that a musician would add less expressive motion to their playing for fear of distracting themselves from the difficult notes. On the other hand, for a slower more lyrical passage, a performer might feel freer to add expressive motions. As outlined in the above scenario, we predicted that performers would move less during a faster, more technically challenging passage than while playing a slower, more lyrical passage. Therefore, we asked the volunteers to bring both a slow lyrical example and a faster more technically challenging example with the following hypothesis: (*H2*): *While adjusting towards more motion in general, participants will prefer a greater amount of performer motion for lyrical pieces than for technical pieces*.

B. Recording Procedure

Each of the four musicians wore a motion capture suit to which small, spherical markers were applied in conventional anatomical locations (Fig.1). Some experimentation was conducted in order to optimize placement of finger markers.

Figure 1. Experimenter in the motion capture suit and markers

Due to their close proximity, fingers are particularly challenging parts of the body to capture using a motion capture system (in this case VICON Blade). The system often confuses the markers, mislabelling fingers and joints. There was also the issue of whether or not wearing the typical motion capture fingerless gloves would interfere with the musicians' ability to play well. To avoid using the gloves and in an attempt to yield a more refined capture of finger motion, we tried to use make-up tape to attach miniature markers directly onto their hands and fingers (Fig.2). However, for most of the musicians, these failed to stay attached during playing. Therefore, after testing to be sure that the gloves minimally interfered with playing level, the gloves were used for the capture including two extra markers taped to the index finger and the minimus (pinky) fingers (Fig.2). Markers were also attached to various points on each instrument to allow for possibility of adding instruments to the animations.

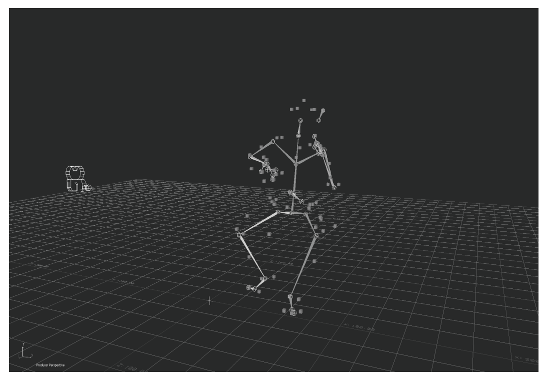
C. Method of Animation

The data for each of these captures were cleaned and imported into MotionBuilder (Catalogue, n.d.) for animation (Fig.3). From the originals, a diminished motion version and an augmented motion version were created. This was achieved by hand-editing what was deemed expressive motion.

Figure 2. Examples of marking the musicians' hands

Expressive motion was operationalized as any additional motion that was not a direct result of technical movements. For example, the cellist tended to sway their upper body from side to side and backwards and forwards. These motions were not direct results of the bowing or fingering motions required

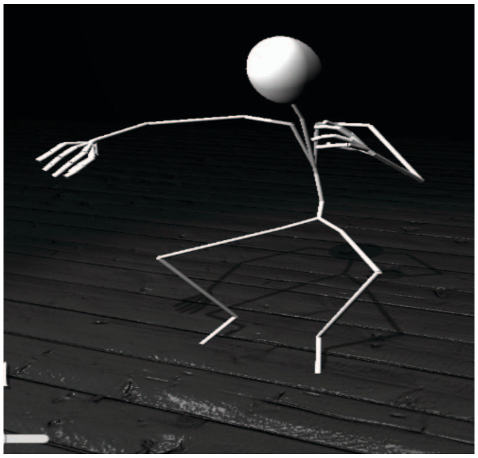
Figure 3. Motion capture data realized in MotionBuilder



to produce notes. Henceforth, this motion is referred to as an expressive motion. For the augmented condition, these same motions were amplified. For the diminished condition, an attempted was made to keep the spine as stationary as possible. A similar process applied in altering the expressive motions of each of the performers. The most commonly altered motions included head motions, knee bends, and full-body swaying.

After creating these animated versions, each animated figure was fitted with white cylinders in a skeletal fashion and placed on a wooden floor with a dark, neutral background (Fig.4). A skeleton model was chosen after attempts to fit more realistic models were found to reduce the apparent motions overall since some body parts were visually obscured. Additionally, part of what made motion capture an ideal tool for this study was that it eliminated all other physical influences, such as gender and attractiveness. A skeletal stick figure animation is devoid of all these other visual aspects. Additionally, there were no chairs or instruments animated into the scenes. This decision was made based on the sheer difficulty of attempting to align the instruments with the live capture data without having the instruments collide with a limb or the main body. It was especially difficult in the augmented and diminished animations because the instrument motion data did not alter

Figure 4. Animated skeletal representation of motion data



naturally when the body motions were altered. The same issues presented themselves when we attempted to add chairs. Therefore, it was deemed too difficult a task and the instruments and chairs were consequently omitted. Participants reported no aversion to the stick figures or to the absence of instruments or chairs when asked during post-experiment interviews.

D. Interface

The interface was created using Unity (Goldstone, 2009). The three versions of each capture (augmented, original, and diminished) were combined into one dynamic video that was adjustable by a slider. From left, to center, to right, the slider continuously altered the live animation from the original motion at the center to mostly the diminished version if moved to the far left and to mostly the augmented version if moved to the far right. This gave the illusion that participants were able to adjust the amount of expressive motion in each video clip in real time without manipulating the audio of the performance.

III. METHODOLOGY

A. Participants

There were 16 participants from the School of Music at Ohio State University. Participants ranged in age between 18-22.

B. Procedure

In order to later test whether musical experience might have an effect on the results, the first part of the interface directed participants to complete the Ollen Musical Sophistication Index (OMSI) and record their results (Ollen, 2006). No *a priori* hypothesis was made regarding possible effects of musical sophistication. Recall that the aim of this study is to test whether musicians with more motion are perceived as better musicians. Therefore, to test this concept with the method of adjustment, we presented participants with the following scenario via these instructions: *"You are going to see eight different performances. For each performance, we want you to consider the following scenario. Suppose there is a performer who is preparing an animated video to submit for a competition. While viewing their video, the performer is thinking that perhaps there is room to 'cheat' a bit by enhancing their performance by manipulating the movement. With the slider at the bottom of each video clip, you can manipulate the overall amount of movement in the performance (left diminishes the motion while right augments the motion). For each animated clip, we want you to tune the slider to give what you think is the most musically superior performance."*

IV. RESULTS

Recall that our first hypothesis predicted that participants would adjust the overall amount of performer motion towards the augmented version. Given that the slider moved from diminished (0) to original (0.5) to augmented (1), in order to reject the null hypothesis, the mean must be significantly higher than 0.5. Consistent with our hypothesis, a single-sample t-test showed that participants chose, on average, to move the slider towards the augmented version of each recording ($M = 0.634$, $SD = 0.218$, $t\ (127) = 6.944$, $p <$ .0001).

Our second hypothesis predicted that participants would prefer more motion for lyrical passages than for technical passages. In fact, using an independent means t-test, we found no difference between the two conditions ($t\ (126) = -0.615$, p = 0.54).

Additionally, we were interested in whether the degree of musical sophistication of the participants affected the results. Using the OMSI index, participants were categorized into two groups; more musically sophisticated (>500) and less musically sophisticated. An independent means t-test demonstrated that there was no statistically significant difference between the two groups of participants ($t\ (126) =$ -1.603, $p = 0.112$).

Finally, a one-way ANOVA determined that there was a significant difference between the average amounts of motion preferred for each of the instruments ($F\ (3, 124) = 5.04$, $p =$ .002). A Tukey *post-hoc* test revealed that the average selected amount of motion for cello ($M = 0.728$, $SD = 0.109$) was significantly higher than the average selected for clarinet ($M = 0.556$, $SD = 0.18$, $p = .007$) or for violin ($M = 0.573$, SD $= 0.246$, $p = .019$). However, it was not significantly higher than flute ($M = 0.679$, $SD = 0.263$, $p = .782$).

V. DISCUSSION

The results were consistent with our first hypothesis in that when participants were instructed to adjust the overall amount of performer motion to enhance the animated musical performances, they tended to augment the overall amount of motion. These results corroborate findings by Morrison *et al.* (2009) and Juchniewicz (2008). Our second hypothesis predicted that participants would select a greater overall amount of performer motion for lyrical passages compared with technical passages. However, this prediction was not born out in the data.

In comparing the mean adjustments for each of the instruments, it was notable that the motion of the cellist was augmented significantly more than that of the clarinet and the violin. There was no statistical difference in comparison to the adjustments for flute. It is important to acknowledge that the expressive motions of each of the instruments differ from each other greatly. Therefore, it is hard to draw any specific conclusions from these resulting differences in adjustments.

Finally, the wording of the instructions may have been a possible confound in this experiment. Our participants were instructed to "enhance" the performance. Unfortunately, the word "enhance" could imply meanings such as "amplify" or "increase". These meanings could have led subjects to believe that the task was to increase the motion of the performance rather than make adjustments to improve the performance, thereby biasing the data.

VI. FUTURE RESEARCH

It would be appropriate to repeat this experiment with a change of instructions, particularly avoiding the word "enhancing". Additionally, it would be of interest to explore further the physical visual elements of performance. Motion capture technology in particular provides a useful way to study performer motion without confounding factors such as attractiveness, gender, race, etc. Additionally, animating the motion data allows for manipulation of visual elements without altering the original audio recording of a performance.

Upon closer inspection of how the sliders manipulated the performer motions, there were sometimes marked changes in posture and head placement. For example, the cellist, when augmented, also adopted a more expansive, open posture and a raised head position. Research on posture and eye contact indicate that open postures and good eye contact communicate friendliness, trustworthiness, and greater emotionality (Burgoon, Birk, & Pfau, 1990; Sundaram & Webster, 2000). It would be of interest to further examine how augmented and diminished motions altered the overall physical vocabulary of each performer to understand what individual components of performer motion might be contributing to these results.

ACKNOWLEDGMENT

Our thanks to the Center for Cognitive and Brain Sciences at Ohio State University for funding this research through a summer research grant. Additional thanks to our performers (Evan Lynch, Mark Rudoff, Snow Shen, Ann Stimson), our motion capture operators (Vita Berezina-Blackburn and Lakshika Udakandage), our unity app developer (John Luna), and the head of the Advance Computing Center for the Arts and Design, Maria Palazzi. Finally, we would like to thank members of Cognitive and Systematic Musicology Laboratory at Ohio State University for critical feedback and support.

REFERENCES

Berry, M. (2009). The Importance of Bodily Gesture in Sofia Gubaidulina's Music for Low Strings. *Music Theory Online*, *15*.

Burgoon, J. K., Birk, T., & Pfau, M. (1990). Nonverbal behaviors, persuasion, and credibility. *Human Communication Research*, *17*(1), 140-169

Catalogue, S. Motion Builder: Making Motion Control Programming Easy.

Fisher, G., & Lochhead, J. (2002). Analyzing from the Body. *Theory and Practice*, 37-67.

Goldstone, W. (2009). *Unity game development essentials*. Packt Publishing Ltd.

Gritten, A., & King, E. (Eds.). (2006). *Music and gesture*. Ashgate Publishing, Ltd..

Hatten, R. S. (2004). *Interpreting musical gestures, topics, and tropes: Mozart, Beethoven, Schubert*. Indiana University Press.

Juchniewicz, J. (2008). The influence of physical movement on the perception of musical performance. *Psychology of Music*.

Larson, S. (2012). *Musical Forces: Motion, Metaphor, and Meaning in Music*. Indiana University Press.

McCreless, P. (2006). Anatomy of a gesture: From Davidovsky to Chopin and back. *Approaches to Meaning in Music*, 11-40.

Montague, E. (2012). Instrumental Gesture in Chopin's Étude in A-flat Major, Op. 25, No. 1. *Music Theory Online*, *18*(4).

Morrison, S. J., Price, H. E., Geiger, C. G., & Cornacchio, R. A. (2009). The effect of conductor expressivity on ensemble performance evaluation. *Journal of Research in Music Education*, *57*(1), 37-49.

Ollen, J. E. (2006). *A criterion-related validity test of selected indicators of musical sophistication using expert ratings* (Doctoral dissertation, The Ohio State University).

Schutz, M., & Lipscomb, S. (2007). Hearing gestures, seeing music: Vision influences perceived tone duration. *Perception*, *36*(6), 888.

Sundaram, D. S., & Webster, C. (2000). The role of nonverbal communication in service encounters. *Journal of Services Marketing*, *14*(5), 378-391.

Wapnick, J., Darrow, A. A., Kovacs, J., & Dalrymple, L. (1997). Effects of physical attractiveness on evaluation of vocal performance. *Journal of Research in Music Education*, *45*(3), 470-479.

Wapnick, J., Mazza, J. K., & Darrow, A. A. (1998). Effects of performer attractiveness, stage behavior, and dress on violin performance evaluation. *Journal of Research in Music Education*, *46*(4), 510-521.

Kinematic Features and Auditors' Perception of Expressive Gesture in Romantic Piano Performances

Catherine Massie-Laberge,*[1] Isabelle Cossette,*[2] Marcelo M. Wanderley*[3]

*Music Research Department, Schulich School of Music, CIRMMT, McGill University, Canada

[1]catherine.massie-laberge@mail.mcgill.ca, [2]isabelle.cossette1@mcgill.ca, [3]marcelo.wanderley@mcgill.ca

ABSTRACT

This work seeks to identify the relationship between pianists' use of motion cues (quantity, velocity, regularity) that convey expressive intentions and auditors' evaluation of these cues, as well as between structural parameters and motion cues. Five pianists played an excerpt from a well-learned piece of their Romantic repertoire in four expressive conditions (normal, exaggerated, deadpan, immobile). A motion capture system tracked pianists' movements and a force plate measured the weight distribution. Twenty auditors rated the performances according to motion cues and to the movements of body parts (head, arms, hands, torso, hips). Quantity of movement of the head and hips, mean velocity of the head, and force applied on the piano stool were overall higher in the exaggerated performances than in other conditions. Head quantity of motion yields significant differences between conditions for every pianist. When comparing expressive conditions, quantity of motion was rated as the most important cue for auditors. Arm and head movements received the greatest attention, while hip movements obtained the lowest attention. Velocity and quantity of motion were associated to phrase boundaries, modulations, sound dynamics, and articulation, while regularity of motion was related to similarity in rhythmic patterns. This work led to the identification of pianists' motion parameters that could be associated to expressive intentions and structural features, and that were perceived by auditors.

I. INTRODUCTION

The development of robust methods to analyze pianists' gestural movements requires one to understand the correlation between pianists' use of motion cues that convey expression and the structural parameters of music. To evaluate how auditors perceive these motion cues in musical performance could also orient better the analysis toward specific kinematic features. Acoustic parameters, such as timing, phrasing, and dynamics, how they influence structural communication in performance, and the way auditors relate these parameters to expressive intentions have been extensively studied (e.g., Clarke ,1989; Gabrielsson & Juslin,1996; Gagnon & Peretz, 2003; Sloboda, 1985). Gagnon and Peretz (2003) found that fast tempi were related to expressions of excitement, surprise, potency, while slow tempi were associated with calmness, boredom, and sadness. Moreover, pianists appear to modify articulation and note length in relation to the different metrical positions of the note (Sloboda, 1985).

The musical structure has been also discussed in relation to movements. According to Delalande (1988), Gould's left hand movements in Bach performances suggest a precise analysis of the score. He would emphasize syncope and anticipates phrases by a preparation with a vibrato of the hand, a support on the up-beat (lower point of the hand's trajectory), and finally a release on the strong beat (rise with the hand). Studies were also conducted under different performance conditions (i.e., immobile, deadpan, standard, exaggerated) to understand the correlation between levels of expression, body movements, and musical structure (e.g., Davidson, 1994; Thompson & Luck, 2012; Wanderley, Vines, Middleton, McKay, & Hatch, 2005). Looking at piano performances, Thompson and Luck (2012) demonstrated that quantity of motion was linked to articulation, dynamic markings, and with the piece's climax. Velocity of head movement was positively correlated with key-velocity (Camurri, Mazzarino, Ricchetti, Timmers, & Volpe, 2004). Short phrases in Chopin's preludes presenting the same rhythmical pattern at different moments in the score entailed similar movements by pianists (MacRitchie, Buck & Bailey, 2009). In a study conducted on clarinetists' gestures, Wanderley and colleagues (2005) revealed that there is a discrepancy in timing between the conditions. The immobile condition was performed faster than standard and exaggerated performances, suggesting that motion is associated to the rhythmic structure of phrases. They also showed that performers use less quantity of motion during very active technical passages, and exaggerate their motions during easier ones. Bell motion in clarinet playing also reinforces idiomatic acoustic events at phrase boundaries and at places with harmonic tension (Teixeira, Loureiro, Wanderley, & Yehia, 2014).

Nusseck and Wanderley (2009) have indicated that certain movements are not essential to appreciate a musical performance. Modifying specific kinematic parameters of clarinet performance (e.g., removing movements from torso or arms and hands) does not necessarily affect its expressiveness and understanding (Nusseck & Wanderley, 2009), and results in nearly no effect on observers' judgments of tension, intensity, and fluency of movements. Moreover, every player performed the piece with personal movement velocities and the total amount of motion differed for the individual body parts (e.g., one may use larger arm motions and another more body sway).

The performance conditions were also used to assess auditors' visual perception of different levels of expression. Auditors were better at identifying performers' expressive intentions with visual information than with sound only (Davidson, 1994). The quantity and velocity of head and torso movement sufficed to discriminate between performance conditions (Davidson, 1994), and between emotions (Dahl & Friberg, 2007). The hip swaying motion towards and away from the keyboards was always present in the different performance conditions but the relationship with the musical structure was not clear (Davidson, 2007).

In the present study we propose a new approach to extract kinematic features of piano performances of pieces from the

Romantic repertoire in the performance conditions used by Thompson and Luck (2012). The objectives are to identify the relationship between pianists' use of motion cues (quantity, velocity, regularity) and acoustic parameters that convey expression and information about the structural parameters of music, and to compare these results to auditors' perception of the same cues.

II. EXPERIMENTAL METHODOLOGY

A. Tracking pianists' movements

Five graduate pianist-students (2 females) performed a thirty seconds Romantic excerpt of their repertoire in four different conditions: normal, deadpan, exaggerated, immobile. Deadpan referred to playing with a reduced level of expression, whereas an immobile performance was restricted to the essential movements to produce a normal performance. Each pianist played one of the following excerpts: Chopin 4th Ballade, Brahms 3rd Piano Sonata, Chopin scherzo no. 2, Schumann Fantasy, Medtner Sonata Reminiscenza op. 38. Each condition was repeated three times for a total of 12 performances per pianist. Data were collected with a Qualisys motion capture system. To capture movements, 49 passive reflective markers apposed to the pianists' hands, arms, shoulders, torso, head, and hips were filmed with ten infrared cameras at a sampling rate of 240 frames per second. The placement of markers is shown in Figure 1.

Force applied on the piano stool was measured with a Bertec force plate. Performances were video recorded with a Sony Wide Angle video camera and audio recorded with a Sennheiser MKH microphone. A Qualisys USB acquisition board was used to capture force plate signals together with the motion capture. The beginning of each measurement was time stamped (SMPTE timecode) at 25 Hz, and a Rosendahl Nanosyncs HD word clock, sampled at 48 kHz, generated the same signal for all the digital devices.

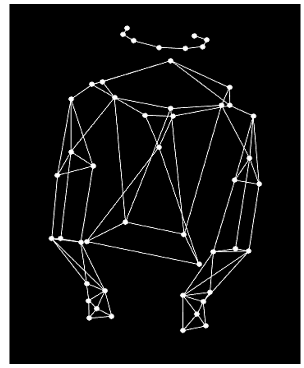


Figure 1. Marker location data for upper body of pianist

Time coordinates of the data were aligned to their corresponding musical event using a time-warping algorithm (Verron, 2005). The exact time of each important gestural event was identified and annotated with Variations Audio Timeliner software. Each piece was segmented into musical notes or phrases and analyzed (i.e., harmony, phrase structure, melody, sound dynamics, and articulation).

B. Auditors' perception of expressive parameters

Twenty auditors (10 females), with at least 5 years of musical training, were asked to associate the musical excerpts to the conditions. They rated the importance of motion cues (quantity, regularity, velocity) on a five-point Likert scale, while being provided with video, or audio and video. They indicated which body parts (i.e., head, torso, shoulders, arms, hands, hips) contributed the most to help them discriminate the expressions. Participants listened to the excerpts as many times as they wanted. Changes in quantity, velocity, periodicity, and force of motion were calculated to compare auditors' answers to the quantitative measurements of pianists' movements, and demonstrate whether there exists a correlation between auditors' perception and pianists' usage of motion cues.

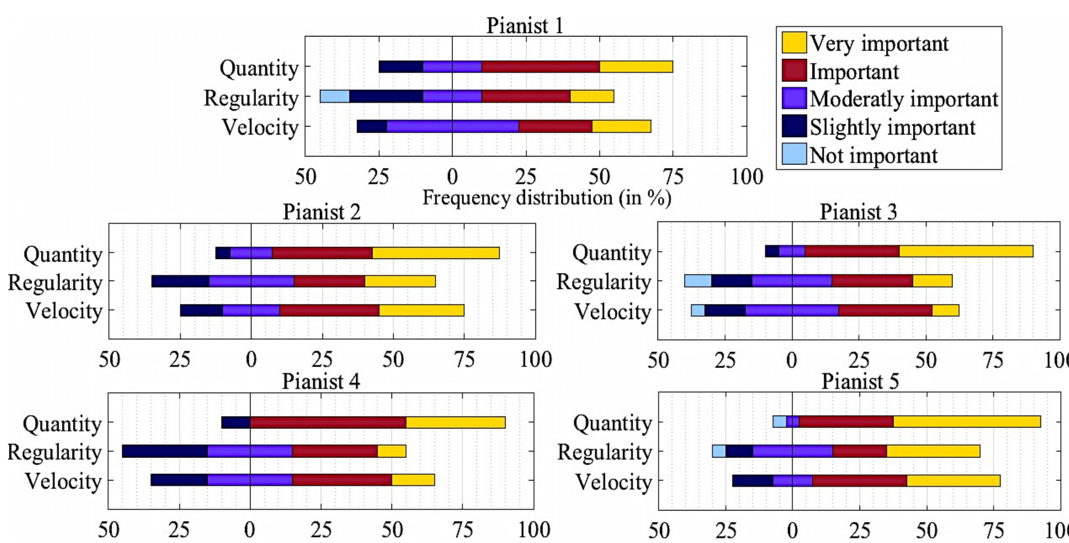


Figure 2. Frequency distribution of auditors' choice of motion cues per pianist

III. RESULTS

A. Auditors' choice of motion cues

The importance and effectiveness of quantity, regularity and velocity of motion cues to convey expressive conditions as perceived by the auditors is shown in Figure 2. A clear pattern can be discerned in auditors' answers. The quantity of motion (mean: 42% very important) was rated as the most important motion cue for every pianist to discriminate between the conditions. It was followed by velocity (mean: 22% very important). The regularity (periodicity) of movement was rated the less important motion cue. Since quantity of motion was perceived by auditors to be the most significant motion cues to decipher the expressive conditions, greater emphasis was put on the analysis of the total distance travelled by different body parts.

B. Evaluating quantity of motion across conditions

To assess whether there are significant differences in the total amount of movement between conditions, we conducted a series of one-way ANOVAs on the head, torso, shoulders, arms, hands, and hips. Each excerpt was segmented according to the phrasing structure.

Table 1. Results of one-way ANOVAs and pair-wise comparison (Tuker-Kramer) on total distance travelled by different body parts.

		Head	Torso	Left shoulder	Right shoulder	Left arm	Right arm	Left hand	Right hand	Hips
P1	ANOVA p-value	*	-	-	*	-	-	-	-	**
	normal-deadpan	-	-	-	-	-	-	-	-	-
	normal-exaggerated	-	-	-	-	-	-	-	-	-
	normal-immobile	-	-	-	-	-	-	-	-	-
	deadpan-exaggerated	*	-	-	-	-	-	-	-	*
	deadpan-immobile		-	-	-	-	-	-	-	-
	exaggerated-immobile	*	-	-	*	-	-	-	-	**
		Head	Torso	Left shoulder	Right shoulder	Left arm	Right arm	Left hand	Right hand	Hips
P2	ANOVA p-value	**	**	**	**	-	-	-	-	**
	normal-deadpan	**	**	*	*	-	-	-	-	*
	normal-exaggerated	-	-	-	-	-	-	-	-	-
	normal-immobile	*	-	-	-	-	-	-	-	-
	deadpan-exaggerated	**	**	**	**	-	-	-	-	**
	deadpan-immobile	-	-	-	-	-	-	-	-	-
	exaggerated-immobile	**	*	*	*	-	-	-	-	-
		Head	Torso	Left shoulder	Right shoulder	Left arm	Right arm	Left hand	Right hand	Hips
P3	ANOVA p-value	*	*	*	-	-	-	-	-	-
	normal-deadpan	-	-	-	-	-	-	-	-	-
	normal-exaggerated	-	-	-	-	-	-	-	-	-
	normal-immobile	-	-	-	-	-	-	-	-	-
	deadpan-exaggerated	-	-	-	-	-	-	-	-	-
	deadpan-immobile	-	-	-	-	-	-	-	-	-
	exaggerated-immobile	*	-	-	-	-	-	-	-	-
		Head	Torso	Left shoulder	Right shoulder	Left arm	Right arm	Left hand	Right hand	Hips
P4	ANOVA p-value	*	*	*	*	-	-	-	-	**
	normal-deadpan	-	-	-	-	-	-	-	-	-
	normal-exaggerated	-	-	-	-	-	-	-	-	-
	normal-immobile	-	-	-	-	-	-	-	-	-
	deadpan-exaggerated	-	*	-	-	-	-	-	-	*
	deadpan-immobile	-	-	-	-	-	-	-	-	-
	exaggerated-immobile	*	*	-	*	-	-	-	-	*
		Head	Torso	Left shoulder	Right shoulder	Left arm	Right arm	Left hand	Right hand	Hips
P5	ANOVA p-value	**	**	**	**	-	**	-	*	**
	normal-deadpan	**	**	**	**	-	*	-	-	**
	normal-exaggerated	-	-	-	-	-	-	-	-	-
	normal-immobile	**	**	**	**	-	-	-	-	**
	deadpan-exaggerated	**	**	**	**	-	*	-	*	**
	deadpan-immobile	-	-	-	-	-	-	-	-	-
	exaggerated-immobile	**	**	**	**	-	-	-	-	**

* $p < 0.05$ ** $p < 0.01$

The results from this analysis are summarized in Table 1. For each pianist, the top row indicates the overall significance level of each ANOVA, while the following six rows designate pair-wise comparisons (Tukey-Kramer) significant at p .05. Major differences occur between the deadpan or immobile performances and exaggerated performances. Difference in quantity of motion of the head between the conditions is significant for every pianist, and the same applies for each pianist's hip motion except for pianist 3. The sole pianist who exhibits significant differences in the right hand motion across conditions is pianist 5.

Figure 3 represents the frequency distribution of auditors' choice of body parts per pianist. The importance given to certain body parts diverges from one performer to another. Overall, we observe that arms (mean: 66%) and head (mean: 47%) were the most frequently chosen body parts that auditors associated to the expressive conditions of the excerpts. The hips receive the lowest attention for every pianist (mean: 8%) except for pianist 4. When the head motion is significantly different across conditions, the same also applies for the torso and shoulders.

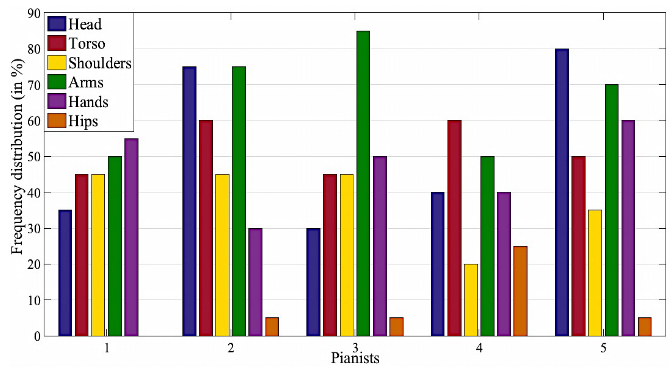


Figure 3. Frequency distribution (in %) of the body parts per pianist that were perceived by auditors as the most helpful to discriminate expression

Although results from Table 1 and Figure 3 do not systematically correspond, both indicate that pianist 3's hip motion was not significant. The arm motion of pianist 3 had a major influence in the discrimination of the conditions, compared to other body parts. Table 1, however, reveals no significant differences between the conditions for arm motion. The absence of movement from the hips throughout the conditions for pianist 3 may explain why auditors focused their attention on the arms and hands. Table 1 reveals greater differences across performance conditions for motion of the head and torso in pianists 2 and 5 than in other pianists. Indeed, the head was chosen respectively 75% and 80% of the time for pianists 2 and 5. We can conclude that the more the level of expression increases, the greater the divergence in the parts of the body that are more distant from the keyboard.

C. Correlation between motion cues and musical structure

Since the head motion of each pianist varied significantly between conditions, structural parameters were first analyzed in relation to the motion of the head. Mean velocity of head motion is the highest in the exaggerated condition for every pianist except for pianist 5. The greatest mean percentage difference (83.22%) occurs between the exaggerated and deadpan conditions for pianist 2, which indicates that this pianist used a wider range of speeds and larger directional changes between the conditions. Figure 4 presents the head velocity pattern for pianist 2 during normal and exaggerated conditions. The velocity patterns seem to follow the section boundaries identified in the score. Indeed, changes occur at the beginning and ending of new sections, and change rapidly at places with modulations. Toward the end of the excerpt, a French sixth chord, successive accentuated chords in a high dynamic level, and a final *rallentendo* are heard.

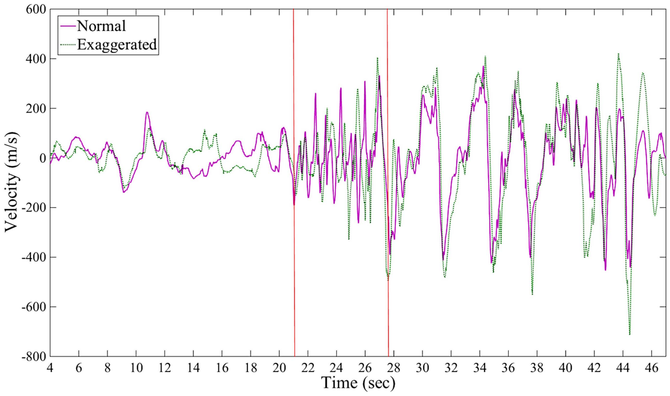


Figure 4. Brahms 3rd Piano Sonata – Head velocity for pianist 2. The red lines indicate section boundaries

As seen in Table 1, the hip motion presented significant differences across the expressive conditions, mainly in pianist 5. Figure 5 illustrates the significant discrepancies observed in the force exerted on the piano stool by pianist 6 during normal and deadpan performances, as well as between normal and immobile performances. The normal and exaggerated performances expose similar amplitude and show two moments with more variations in the force. The first one corresponds to an accentuated dominant chord with *forte* dynamic followed by a dominant seventh chord (between 21 s and 25 s), and the second one occurs at the climax of the piece (at 33.5 s). We also found peaks in force activity before accentuated chords and at the section boundaries.

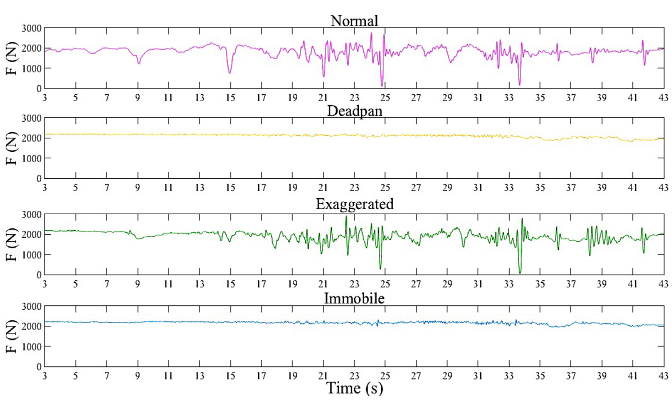


Figure 5. Medtner Sonata Reminiscenza – Force applied on the piano stool by pianist 5

Figure 6 exposes pianist 3's motion of the head in the four expressive conditions. A short four note motive in the low register with an accentuated last note, which is reiterated throughout the excerpt, is characterized by a periodic motion of the head, mainly in the exaggerated and normal conditions. The more the level of expression increases, the further away from the keyboard the head goes. In addition, changes in velocity are related to high dynamic level, large chords in the low register, and a staccato articulation (e.g., between 15 and 17 s).

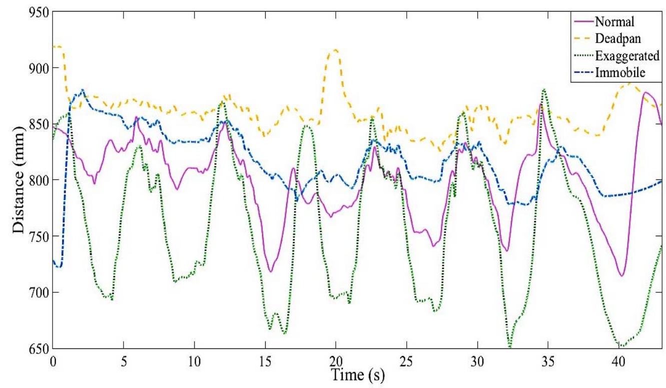


Figure 6. Chopin scherzo no. 2. – Pianist's 3 head position data

In the more rhythmical section (between 20 and 26 s), the head has less frequent changes in velocity due to a dotted rhythm and bigger chords to reach. In this section, the complex rhythm does not leave place for exaggeration, since the left and right hands have greater intervals to reach.

IV. DISCUSSION

In this article, we have exposed the relationships between pianists' physical gestures and their efficiency in conveying expression and structural parameters of pieces from the Romantic period. Furthermore, we compared the gestural data to auditors' subjective perception of different expressive conditions. The analysis of kinematic features confirms the relationship between movements and musical structure. We also established that auditors, while associating the excerpts to the expressive conditions, concentrate more on quantity of motion than velocity or regularity of motion, and that these cues are most noticeable in the arms, head, and torso. Individual differences still occurred between pianists, which auditors noticed by focusing on different cues and body parts depending on the pianist.

The deadpan performances were played with less quantity of motion than other conditions. Quantity of motion from the head, torso, and hips yield significant differences mainly between the exaggerated and immobile conditions. Pianist 5's right hand yields significant differences between conditions. This suggests that, for some pianists, parts of the body close to the instrument may have an expressive and structural function. The swaying motion from the hips was not always present in every performance condition nor was it present for all pianists, which differs from Davidson's findings. It was significantly reduced in the immobile condition, and even more in the deadpan condition. Results obtained from the force plate measurements have revealed a clear connection between motion from the hips and structural parameters of music. Moreover, Davidson (1994) observed that the pianists' hand movements provided little information about the expressiveness of the performer. Conversely, this study has showed a significant difference in the quantity of movement between the performing conditions in the movements of the wrists and hands.

Studies have pointed out a few relationships between the phrasing structure, melodic line, and performers' gestures (Camurri et al. 2004; MacRitchie et al. 2009; Teixeira et al. 2014; Thompson & Luck, 2012; Wanderley et al., 2005). This exploratory work considered a large-scale analysis (harmony, phrase structure, melody (motive), sound dynamic, and articulation), and investigated various piano pieces with different levels of complexity. Similarly to Thompson and Luck's results, quantity of motion was mainly associated to phrase boundaries, motives, modulations, and repetitions. Velocity of motion was related to staccato articulation, sound dynamics, and chords. Periodicity of the head motion was found in similar rhythmic patterns. Force applied on the piano stool was associated to section boundaries, and before accentuated chords. The recurrence of these motion cues in different musical contexts reinforces the theory that certain gestures effectively support musical structural parameters.

The diversity of repertoire helped us make various relationships between the compositional structure and gestures. Future work is needed to characterize the influence of these motion cues on auditors regarding the musical structure. A better understanding of the impact of individual motion cues on the appreciation and comprehension of the musical structure could further expand the scope of use and validity of this approach. The method would need to be reiterated by having each pianist play the same Romantic excerpts in order to evaluate the reproducibility of gestures in different performances and different piano schools.

This work may also contribute to the development and design of new pedagogical technologies to help students acquire movement awareness and increase their musical communicative abilities.

ACKNOWLEDGMENT

We gratefully acknowledge financial support from the Social Sciences and Humanities Research Council of Canada. We would also like to thank all the performers and auditors for their participation and collaboration.

REFERENCES

Camurri, A., Mazzarino, B., Ricchetti, M., Timmers, R., & Volpe, G. (2003). Multimodal analysis of expressive gesture in music and dance performances. In A. Camurri & G. Volpe (Eds.), *Gesture-based Communication in Human-Computer Interaction* (pp. 20-39). Berlin: Springer Verlag.

Clarke, E. F. (1987). Levels of structure in the organization of musical time. *Contemporary Music Review*, *2*(1), 211-238.

Dahl, S., & Friberg, A. (2007). Visual perception of expressiveness in musicians' body movements. *Music Perception: An Interdisciplinary Journal*, *24*(5), 433-454.

Davidson, J. W. (1994). What type of information is conveyed in the body movements of solo musician performers. *Journal of Human Movement Studies*, *6*, 279-301.

Davidson, J. W. (2007). Qualitative insights into the use of expressive body movement in solo piano performance: A case study approach. *Psychology of Music*, *35*(3), 381-401.

Delalande, F. (1988). La gestique de Gould: éléments pour une sémiologie du geste musical. In G. Guertin (Ed.), *Glenn Gould, Pluriel* (pp. 83-111). Montréal: Louise Courteau éditrice.

Gabrielsson, A., & Juslin, P. N. (1996). Emotional expression in music performance: Between the performer's intention and the listener's experience. *Psychology of Music*, *24*(1), 68-91.

Gagnon, L., & Peretz, I. (2003). Mode and tempo relative contributions to "happy-sad" judgements in equitone melodies. *Cognition & Emotion*, *17*(1), 25-40.

MacRitchie, J., Buck, B., & Bailey, N. J. (2009). Visualising Musical Structure Through Performance Gesture. In *Proceedings of the 10th International Society for Music Information Retrieval Conference* (pp. 237-242). Kobe, Japan.

Nusseck, M., & Wanderley, M. M. (2009). Music and Motion—How Music-Related Ancillary Body Movements Contribute to the Experience of Music. *Music Perception: An Interdisciplinary Journal*, *26*(4), 335-353.

Teixeira, E. C., Loureiro, M. A., Wanderley, M. M., & Yehia, H. C. (2015). Motion analysis of clarinet performers. *Journal of New Music Research*, *44*(2), 97-111.

Thompson, M. R., & Luck, G. (2012). Exploring relationships between pianists' body movements, their expressive intentions, and structural elements of the music. *Musicae Scientiae*, *16*(1), 19-40.

Verron, C. (2005). *Traitement et visualisation de données gesturalles captées par Optotrak* [*Processing and visualizing gesture data captured by Optotrak*]. Unpublished manuscript, Input Devices and Music Interaction Laboratory, McGill University, Montréal, Canada.

Wanderley, M. M., Vines, B. W., Middleton, N., McKay, C., & Hatch, W. (2005). The musical significance of clarinetists' ancillary gestures: An exploration of the field. *Journal of New Music Research*, *34*(1), 97-113.

Effects of Temporal Congruence on Evaluations of Conductor Efficacy

Cory D. Meals[1], Steven J. Morrison[2], Deborah Confredo[3]
[1]School of Music, Kennesaw State University, Kennesaw, Georgia, USA
[2]School of Music, University of Washington, Seattle, Washington, USA
[3]Boyer College of Music and Dance, Temple University, Philadelphia, Pennsylvania, USA

Instrumental ensemble conductors are seen as both interpreters of musical information and coordinators of musical activity. The qualities of their gestural contributions have been shown to influence observer evaluations of ensemble performance, but less empirical work exists exploring the influence of temporal action/sound congruence in this setting.

The primary objective of this study is to investigate the role of action/sound congruence in observer evaluations of conductor efficacy where domain specific temporal distance has been experimentally manipulated.

We collected video from five conductors and extracted two excerpts per conductor (fast, slow). These were manipulated using video editing software to create stimuli encompassing intact, audio-lead, and video-lead conditions of ±15% and ±30% of performed tempo. Stimuli were ordered so that no two conductors or conditions (a/v offset or fast/slow excerpt) were seen consecutively. For each excerpt participants (n=110) evaluated the conductor, ensemble, and overall performance.

We compared standardized participant ratings for conductor, ensemble, and overall performance using multivariate ANOVA. Ensemble and conductor ratings across all offset conditions were lower than those for intact stimuli. Significant differences in conductor ratings were found across all offset conditions, most notably +30% video-lead ($F(1,109) = 33.75$, $p < 0.01$) with a moderate effect (partial $\eta^2 = 0.24$). Significant ($p < 0.01$) differences were also found in ensemble ratings of both +15% and -15% offset conditions.

These results suggest that gesture's temporal distance from its sonic correlate plays a detrimental role in observer evaluations of conductor efficacy. While lower ratings of audio-lead offset conditions lend support to a predictive or delineative function of movement in music performance, the presence of similar, negative ratings in video-lead conditions tempers any universality that might be attributed to the aforementioned function. These results support existing research on the multimodal influence of gesture on perception of sound, but highlight an underexamined aspect of action/sound congruence and its influence on aesthetic evaluations.

Measuring when we breathe with music

Finn Upham[1]
[1]Music and Performing Arts Professions, New York University, New York City, NY, USA

Introduction

What can we learn about someone's experience of music by tracking their breathing? Potentially a great deal. At any moment, how we are breathing is determined by what we are doing, what we are thinking, how we are feeling, and who we are. Beyond changes in respiration rate a few studies have found instances of listeners' breathing aligning to particular moments of musical stimuli. Measuring when and how participants show significant respiratory adaptation may open new avenues to capturing important elements of their responses to music as it plays.

Aims

Respiratory alignment to music can but does not always occur. This research presents a method of evaluating disruptions in a passive listener's breathing which may be support adaptation to the music heard, and quantifying the likelihood of stimulus influence in those cases when moments of alignment are identified from other listenings.

Method

Alignment between multiple respiratory sequences to a given musical stimulus, or the absence there of, is an indicator of whether this music encourages alignment. From continuous chest expansion measure from individual participants to many listenings of a variety of pieces, the respiratory sequences are classified as showing none, some, or strong indications of adaptation. Disruptions and changes in breathing patterns are quantified and statistics on their qualities and occurrences are compared between these classes.

Results

In the respiratory sequences measured during a set of repeated response case studies, twelve listening sessions to eleven stimuli from four participants with distinct musical histories, preliminary results show evidence of adaptation. These data show promising variation in each participant's sensitivity to specific stimuli, ranging from a Beethoven string quartet to Maria Carey, and even between single listenings. Possible correlations with ratings of liking, focus, and familiarity will be discussed in more detail.

Validation of a Simple Tool to Characterize Postural Stability in the Setting of the Indian Classical Dance form, *Bharatnatyam*

Nandita K Shankar[1], Ramesh Balasubramaniam[2]
[1]Cell and Molecular Biology, Tulane University, New Orleans, LA, USA
[2]Cognitive & Information Sciences, University of California, Merced, CA, USA

Background
Upright postural control is pivotal to the integrity of dance movement. Most techniques to assess postural stability require expensive instrumentation, not readily accessible to dance schools, especially in countries such as India. This is a particularly striking gap, given the rich heritage of Indian classical dance.

Aims
We sought to evaluate the simple Star Excursion Balance Test (SEBT) as a tool for assessment of postural stability in the context of the *Araimandi*, (half-sitting posture) a primary and fundamental position in the Indian classical dance form, *Bharatanatyam*.

Methods
Twenty-two female volunteers (one highly trained and experienced dancer as index subject and 21 dancers-in-training) underwent the SEBT procedure and a digital still motion capture recording during the *Araimandi* stance. SEBT scores and joint angles computed using MATLAB ™ software. In the index subject alone, postural stability in a standardized optimal *Araimandi* position was also assessed using a laboratory force platform. Joint angle data from each volunteer were compared against those from the index subject force plate analysis, to define degree of alignment or deviation from the index subject. We evaluated the relationship between joint angle data and SEBT scores.

Results
Subject characteristics (mean, range): age: 20 yrs. (7-42), BMI: 20 kg/m^2 (14-29); 4 subjects had no additional physical activity. Preliminary data analysis reveals strong differences between the expert and novice performers.

Conclusion
To our knowledge, this is the first experiment to attempt to develop a simple method to quantitatively characterize postural stability in the context of the Indian classical dance form, *Bharatanatyam*. If the SEBT score is indeed a viable tool for assessment of postural stability in such a setting, it could be readily applied in dance schools, and perhaps other settings such as rehabilitation centers in developing nations with limited resources.

Singing about sharing: Music making and preschoolers' prosocial behavior

Sara Lynn Beck[1] & John Rieser[1]
[1]Psychology & Human Development, Vanderbilt University, Nashville, TN, USA

Educational children's media relies heavily upon musical sequences to convey prosocial messages about desirable behaviors, but little is known about the conditions under which children's social behavior is most likely to be influenced by engaging with music through movement or singing. Recent research has shown that engaging in joint music making can result in increased cooperative behavior among preschoolers (Kirschner & Tomasello, 2010) and that moving to a musical beat can foster increased prosocial behavior in infants as young as 14 months (Cirelli, Einarson, & Trainor, 2014). No studies to date have examined the impact of prosocial lyrics on children's behavior. The current study represents a starting point for examining the unique contribution of lyrics and interpersonal synchrony to children's behavior in the context of music making. We investigated 4 and 5-year-olds' sharing behaviors following one of four experiences: engaging in joint singing and movement to a novel song with either prosocial or neutral lyrics, and engaging in joint non-musical play involving a novel rhyme with either prosocial or neutral content. The four conditions utilize identical props and matched interaction with the experimenter and a research assistant. Children's sharing behavior was coded, and a *t* test on the first two conditions collected (music/prosocial lyrics and play/neutral lyrics) shows a trend toward significance such that children in the music/prosocial condition share more readily than those in the play/neutral condition ($t(18)=-1.907$, $p=.07$). Changes in children's affective state (indicated by each participant on a pictorial likert scale) do not account for the difference between the two groups, suggesting that the difference is not mediated by affect. Data collection and analysis is underway for the other two conditions, which will enable us to begin to examine the unique contribution of prosocial lyrics and interpersonal synchrony in facilitating prosocial behavior.

Measuring Stress in Cases of Discrepancy Between a Task Description and the Accompanying Gesture, Using the Example of Conducting

Sarah Platte,[*1] Rosalind Picard,[*2] Morten Schuldt-Jensen[#3]

[*] *MIT Media Lab, Massachusetts Institute of Technology, Cambridge, MA, USA*

[#] *Department of Choral and Orchestral Conducting, University of Music, Freiburg, Germany*

[1]platte@media.mit.edu, [2]picard@media.mit.edu, [3]schuldtjensen@gmail.com

ABSTRACT

Professional musicians face a wide range of occupational health problems ranging from instrument-related injuries to mental stress. Many attempts to improve the health of musicians have been made during the last decade, mainly by inducing changes in labor conditions (better chairs, hearing protection, etc.), but an important factor has been neglected so far: The influence of the conductor's gestures and rehearsal procedure on the musician's physiology. Our study investigates differences in stress-levels in cases of contrasting and coherent gesture-task-combinations in the communication between conductor and musician.

The within-subject experiment included 19 participants. During the experiment they were instructed to follow videos of varying conducting gestures by pressing a touch-sensor. We used gestures showing differences within two musical categories – dynamics and articulation. Every tapping-event induced an audio feedback mirroring the length and volume of the evoked tone. The length of each tapping event and the amount of pressure was recorded. While following the protocol of coherent and contrasting tasks we also recorded heart rate variability, electrodermal activity and breathing.

The results of the experiment reveal a significant relationship between contrasting gesture-task-combinations and increased physiological stress as measured from heart rate variability. Furthermore, sound quality resulting from incoherent instruction/gesture-combinations deviated negatively from the results of the coherent combination.

By showing that conducting-gestures have an actual and direct influence on the musician's physiology and that a mismatch between a conductor's verbal demand for a specific sound envelope shape and the concurrent conducting gesture has a negative impact on both the musician's stress-level and the resulting sound itself, this study adds an important factor to the knowledge about occupational illness in a vulnerable group of professionals.

I. INTRODUCTION

For many years researchers within the fields of Music Physiology and Musicians' Medicine have been exploring health effects of the instrumental and singing practice of professional musicians (Steinmetz, 2015). Due to the research focus, occupational illness among musicians is mostly seen as related to factors like overstraining, poor playing technique and bad posture, general working conditions (no ergonomic chairs, insufficient light, noise exposure etc.) and psychosocial ambience conditions, but an important factor has been disregarded so far – the impact of conducting-gestures and the accompanying verbal instructions – and the combination of them – on the physical playing actions and stress level of ensemble musicians (Richter et al., 2011).

String players of professional symphony orchestras report conspicuous (preventative) medical leaves in the group, when certain conductors come to work with the orchestra, allegedly because they have experienced that the specific gestural repertoire and rehearsal practice of these conductors provoke tendinitis. The same picture holds for professional choirs, where certain conductors have the dubious reputation of wearing the singers vocally off with their gestures. Based on such reports it seems reasonable to hypothesize that different conducting gestures can have an impact, negative or positive, on the musician's health.

Besides the mere gestures, verbal instructions constitute a major part of the rehearsal procedure. We hypothesize that not only certain conducting movements but also a discrepancy between verbal instruction and shown gesture would evoke negative stress, which – over the long term – could result in stress-related occupational illnesses.

The relative effectiveness of verbal instructions and conducting gestures in a high school choral context, as investigated by Jessica Napoles (2014), is particularly relevant for our experiment design. Napoles asked 44 high school students (Grade 9–12) to sing the song "Music Alone Shall Live", while viewing a video with conducting gestures and following different types of tasks regarding word-stress and articulation: (a) Follow the conducting gesture; (b) Contrasting task; (c) Coherent task. The resulting audio recordings were evaluated by 30 experienced secondary choral teachers and then statistically analyzed by Napoles. The results showed a significantly higher quality rating of performances under the condition of verbal instructions with consistent gestures. Interestingly, the sounding results of inconsistent gesture-task combinations were rated higher than those of the gesture-only tasks. Unfortunately, besides a brief description, Napoles provides no further information about the exact form/trajectory of the used conducting gestures for the shown musical parameters and the decision process for the selection of the specific beat-patterns.

Based on findings on intuitively evoked reactions of basic conducting patterns (Platte et al., 2016) and in view of Jessica Napoles' (2014) findings on the effectiveness of verbal instructions and conducting gestures respectively, this study explores the impact of different pairings of conducting gestures and verbal instructions on length and volume of the elicited tone and on the musician's stress-level.

II. DATA COLLECTION

The presented study was approved by the Committee on the Use of Humans as Experimental Subjects (COUHES) at the Massachusetts Institute of Technology. Before taking part in the study, all participants were informed of purpose and procedure of the experiment and informed consent was obtained.

Our experiment took place in the Bartos Theatre, a lecture hall with a seating capacity of 190. We decided to use this particular room for the creation of a performance-like atmosphere without external sources of noise to animate the participant to perform the test with a high level of concentration. The experimental setup consisted of a screen, showing videos of conducting gestures, headphones for audio feedback and a touch-sensor (Force Sensing Resistor (FSR)) detecting physical pressure. The touch-sensor was custom-built, using a single zone FSR with an accuracy (force repeatability) of +-2% and a force sensitivity range of ~0.2N – 20N, measuring touch events in Millivolt. The collected data was sent though an Arduino UNO via serial communication to Max/MSP, where it was time stamped and saved into a csv file.

In order to measure current stress levels the participants were asked to wear two electrodes to record electrodermal activity (EDA) and an optical sensor for blood volume pulse (BVP) measurement on the fingertips of the non-dominant hand. Additionally, a chest belt was used to record respiration frequency and depth. All sensors were connected to a FlexComp Infiniti Decoder, which sampled the received signals at 2048 samples/second, digitized, encoded, and transmitted the sampled data via the computer's USB port to the BioGraph Infiniti Developer Tools. The software was used to merge and timestamp the received data, as well as to create one single file per task to be saved as csv-file. Max/MSP was used to start the conducting videos synchronized with the data-collection from the touch sensor and to time stamp and ultimately save the collected data.

The within-participants experiment with two randomized sets included 19 healthy participants of both genders (12 male / 7 female) with an age between 19 and 32 ($\bar{x}$ 26.5). 14 participants were musicians/singers who had already or currently played or sung with a conductor; 5 were non-musicians who had never worked with a conductor.

After mounting the sensors as described above, the individual baseline of the participant was measured during a 3-minute video containing relaxing music and nature photographs. Subsequently the participant was instructed to follow videos of varying conducting gestures by playing a sound elicited by pressing the touch sensor. Every tapping-event induced an audio feedback, the length and force of the participant's pressure mirroring the length and volume of the evoked tone respectively. The on- and offset-time of each tapping event and the amount of pressure on the sensor were recorded. The collected sensor data was synchronized with the recorded data of the shown conducting gestures to allow further investigations.

Based on findings on intuitive reactions of conducting gestures (Platte et al. 2016), we decided to use concave and convex trajectories and different sizes of the convex conducting pattern to design a procedure of both coherent and incoherent tasks, subdivided into two categories:

Category 1 – Coherent task gesture combination

The first category of tasks included coherent tasks and gestures (video shows / verbal instruction):

1a) Concave gesture / 'Play short notes'

1b) Convex gesture / 'Play long notes'

1c) Big gesture / 'Play forte'

1d) Small gesture / 'Play piano'.

Category 2 – Incoherent task gesture combination

For the second category of tasks we used contrasting instructions and gestures:

2a) Convex gesture / 'Play short notes'

2b) Concave gesture / 'Play long notes'

2c) Small gesture / 'Play forte'

2d) Big gesture / 'Play piano'.

We executed every task twice in a randomized sequence to prevent order effects, and we closed the experiment with a short questionnaire, asking for background information on age, gender, profession, and musical education/experience.

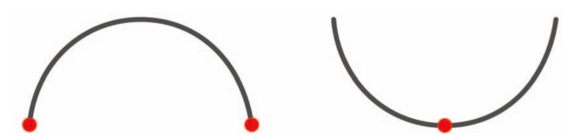


Figure 1. Concave (left), convex (right) conducting patterns with red beat points.

III. RESULTS

Measurements of EDA and breathing did not show statistically significant differences in tasks with predicted higher stress levels. The lack of response in EDA is probably due to the relatively short intervals of the different tasks; longer single tasks with each gesture-task-combination could bring more consistent results. Concerning breathing amplitude, only non-significant trends could be found as discrimination from low-stress to high-stress tasks. However, breathing of other musicians (wind players), and especially singers, could be influenced by stress, as it plays a more "active" part in actual tone production for them than in our one-finger-study.

While on the one hand the heart rate (HR) showed only little difference, on the other hand changes of the heart rate variability (HRV) revealed a strong correlation with higher stress levels. Research has shown a highly significant correlation between short-term HRV analysis and acute mental stress (Tremper 1989, Sinex 1999, Castaldo et al. 2015). Using the collected blood volume pulse (BVP) data, we calculated low frequency (LF | 0.04–0.15 Hz) and high frequency (HF | 0.15–0.4 Hz) powers and used the power ratio of LF/HF as an indicator for mental stress. As HF spectrum changes are associated with changes in mainly vagal-cardiac modulation and changes in LF power with both vagal- and sympathetic-cardiac modulation, a higher LF/HF power ratio indicates a higher level of mental stress. However, only 16 participants showed usably noise-free BVP measurements – data from 3 participants were insufficient and could not be used for calculations of HRV and LF/HF power ratio.

A. Dynamics

Within the musical parameter of dynamics the instruction "play forte" revealed the clearest differences in HRV LF/HF power ratio (p = 0.0003). Almost all participants (15 out of 16) showed higher values of LF/HF power in the cases of discrepant task and shown gesture. However, the amount of evoked stress differs significantly between the groups of musicians and non-musicians (see Figure 2). This is reflected in a higher significance (p=0.003) for musicians, with a mean low-stress value of 1.2 and a ratio of 3.2 for the high-stress condition. Non-musicians showed a similar power ratio (1.1) in cases of low stress, with a less increased high-stress value (2.0) compared to the group of musicians, leading to a higher but still significant p-value of 0.01. Interestingly, in the case of high-stress, the standard deviation of the group of musicians (2.9) is much higher than it is for the group of non-musicians (1.8), indicating that some musicians seem to be accustomed to this type of stress than others.

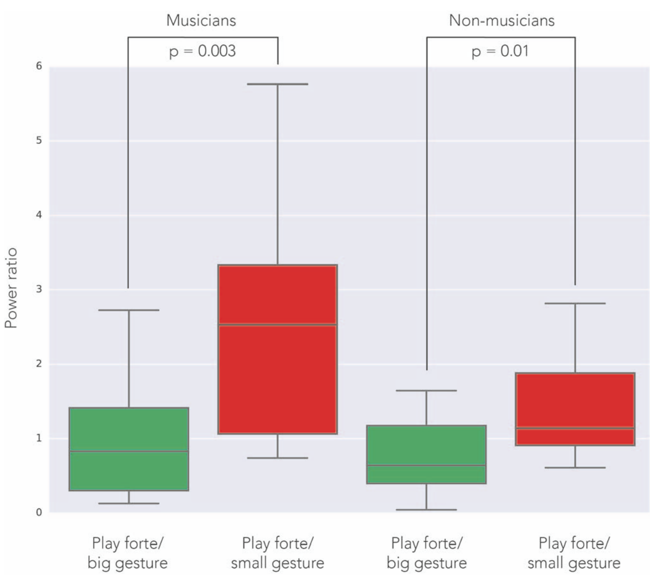


Figure 2. Comparison of HRV LF/HF power ratio in coherent and opposed task-gesture-combinations: play forte | small gesture gesture for high Stress (red) and play forte | big gesture for low Stress (green).

The results of the second task of coherence and incoherence ("play piano") were also significant. Exactly as was the case with the instruction "play forte", the "coherent vs. non-coherent play piano" task also evokes higher differences between low and high stress levels for the group of musicians (low stress 1.2 / high stress 2.2) than for the non-musicians (low stress 1.6 / high stress 1.9), resulting in a significant difference (p=0.002) for our set of 14 musicians, but a non-significant p-value of 0.58 for our smaller set of 5 non-musicians. In contrast to the "play forte" condition, standard deviations of musicians and non-musicians are both high for the high stress task (1.4 / 1.5), while the low stress task shows a very low standard deviation of 0.6 in musicians compared to 1.5 for non-musicians.

B. Articulation

The combination of the instruction "play staccato" while following a convex or a concave conducting gesture, showed the most consistent stress reactions (p = 0.0003). All 16 participants show higher LF/HF power ratios under the stress inducing condition. As for all other tasks, musicians again show a higher significance in the consistency of their stress reaction. As opposed to the tasks of contrasting dynamics, the highest stress reaction in the "play staccato" task is found in the group of non-musicians with a value of 3.2, compared to 2.5 in musicians (see Figure 3). The standard deviation in the case of the low-stress task is very low at 0.6 for both groups, while it is much higher for non-musicians (1.9) compared to 1.2 in musicians for the high-stress condition.

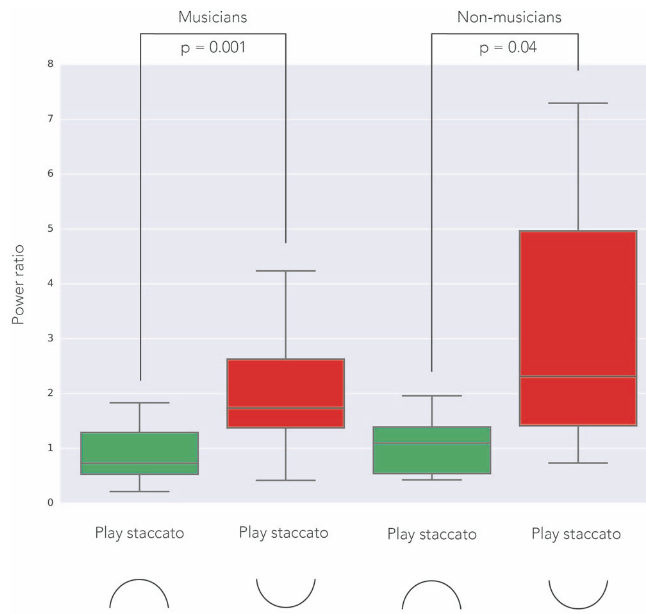


Figure 3. Comparison of HRV LF/HF power ratio in coherent and opposed task-gesture-combinations for musicians and non-musicians: Play staccato | convex gesture for high stress (red) and play staccato | concave gesture for low stress (green).

The "play legato" task also showed a statistically significant finding in change of power ratios, while the differences between coherent and opposed instruction-gesture-combination in this task were smaller than in the previous one (p = 0.001). Furthermore, compared to all other stress inducing tasks, the "play legato" condition evokes the least amount of stress with an overall mean of 2.0. Between the two groups of participants, differences in values and standard deviations are rather small, and can be found only in the results of the stress inducing condition with a power ratio of 2.1 (STD 1.9) for non-musicians and 1.9 (1.2) for musicians.

C. Differences in Length and Amount of Pressure

All resulting reactions of incoherent instruction-gesture-combinations deviated negatively from the results of the corresponding coherent combination. The most striking result can be found in the "play legato" task, showing a highly significant difference (p=0.00003) in the lengths of the evoked pressure on the FSR sensor in the case of a discrepancy between instruction and gesture (see Figure 4). Although both tasks should theoretically elicit the same length of pressure, the participants seemed not to be able (despite instruction) to hold the note as long with the concave gesture (654.3 ms) as with the convex gesture (706.7 ms). As seen in Figure 5, a similar reaction can be found in the “play piano” task, evoking a significantly higher pressure (p=0.03) when accompanied by a big gesture. Neither the “play staccato”, nor the “play forte” instruction elicits significant alterations in the length and amount of pressure respectively, when combined with an incoherent gesture.

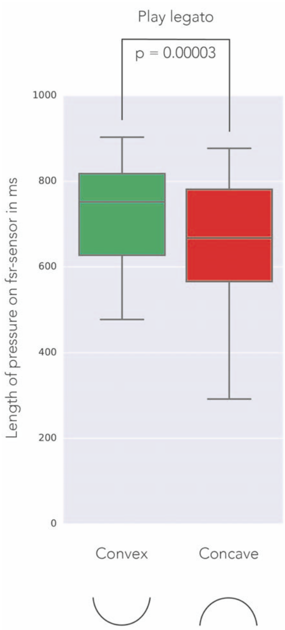


Figure 4. Comparison of length of tone in coherent and opposed task-gesture-combinations for articulation.

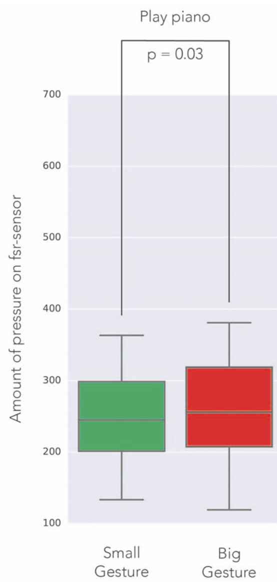


Figure 5. Comparison of pressure in coherent and opposed task-gesture-combinations for dynamics.

IV. DISCUSSION

The analysis of heart rate variability from blood volume pulse data shows a significantly higher stress level in all cases of discrepancies between a given instruction and the accompanying gesture. The negative deviations from the sounding result of the coherent combination, especially in case of articulation, seem to be caused by a relative dominance of the perceived haptic qualities of the gestures, which may be hard or impossible to suppress. The attempt to play exactly as the instruction demands when gestures are incoherent, seems to increase mental stress, while “disobedience” concerning the given verbal instruction seems to be less stressful, as could be seen under the “play legato” condition. Thus, it might be concluded that gestures are more assertive than verbal instructions.

An important implication of these findings is that not only general working conditions but also a lack of consistency between shown gesture and given verbal instruction induces negative stress, with significant cardiac impact, and – over a longer period of time – could potentially contribute to physical and mental illness.

V. CONCLUSION

With this study we demonstrated that a mismatch between the shown conducting gesture and the concurrent verbal demand for a specific sound envelope shape has a negative impact on both the musician’s stress level and the resulting sound itself. Thus, the key to a healthy conducting and rehearsing environment seems to involve supporting a consistency of the conductor’s imaginative musical goal, his or her shown gesture, and the verbal demand.

Recent studies show, that one third of absences at work are due to mental stress and anxiety (Hope, 2013), which in a chronic state is known to lead to serious somatic diseases, such as hypertension, coronary artery disease, heart attack, etc. (Boonnithi & Phongsuphap, 2011). Hence, proving a direct influence of incoherent gesture-task combinations on the musicians’ stress level – and taking it into practical account – could have a major impact on the diagnosis, the treatment and – not least – the prevention of some occupational illnesses among professional musicians. Likewise, a change of rehearsal methods towards more coherence of verbally predetermined musical tasks, conducting gestures, and aimed interpretation could improve not only the daily working environment of the 193,300 professional musicians (in the U.S.), but also lead to a more joyful and healthy experience for innumerable amateur singers and musicians (U.S. Bureau of Labor Statistics, 2015).

REFERENCES

Boonnithi, S., & Phongsuphap, S. (2011). Comparison of heart rate variability measures for mental stress detection. In *Computing in Cardiology, 2011* (pp. 85–88). IEEE. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6164508

Castaldo, R., Melillo, P., Bracale, U., Caserta, M., Triassi, M., & Pecchia, L. (2015). Acute mental stress assessment via short term HRV analysis in healthy adults: A systematic review with meta-analysis. *Biomedical Signal Processing and Control*, *18*, 370–377.

Hope, C. (2013, June 26). One in three absences at work due to anxiety and stress, official Government survey finds. Retrieved from http://www.telegraph.co.uk/news/health/news/10143915/One-in-three-absences-at-work-due-to-anxiety-and-stress-official-Government-survey-finds.html

Napoles, J., jessica. ed. (2014). Verbal Instructions and Conducting Gestures: Examining Two Modes of Communication. *Journal of Music Teacher Education*, *23*(2), 9–20.

Platte, S., & Schuldt-Jensen, M. (2016). Conducting – an Intuitive Gestural Language? Exploring the Actual Impact of Conducting Gestures. In *Proceedings of the 14th International Conference on Music Perception and Cognition*. San Francisco, CA

U.S. Bureau of Labor Statistics: Occupational Outlook Handbook (2015, December 17). Retrieved April 1, 2016, from http://www.bls.gov/ooh/

Richter, B., Zander, M., Hohmann, B., & Spahn, C. (2011). Gehörschutz bei Musikern. (German). *Hearing Protectors in Musicians. (English)*, *59*(6), 538.

Sinex, J. E. (1999). Pulse oximetry: Principles and limitations. *The American Journal of Emergency Medicine*, *17*(1), 59–66.

Steinmetz, A. (2015). Muskuloskeletale Funktionsstörungen bei professionellen Musikern. (German). *Musculoskeletal Dysfunctions in Professional Musicians. (English)*, *53*(1), 13.

Tremper, K. K. (1989). PUlse oximetry. *Chest*, *95*(4), 713–715.

Movement restriction when listening to music impacts positive emotional responses to and liking of pieces of music

Shannon Wright[1], Evan Caldbick[1], and Mario Liotti[1]
[1]Department of Psychology, Simon Fraser University, Burnaby, B.C., Canada

Motor engagement with music occurs frequently in humans at a both a conscious and an automatic level. Groove music is defined as music that elicits a pleasurable response of wanting to move one's body to the music (Janata et al., 2011). The primary aim of this study is to provide further understanding of the impact of body movement on musically-induced emotional responses. Focusing on positively-valenced emotional responses to music, the impact of three types of movement on musically-induced emotion and liking of musical pieces was investigated. Twenty-seven subjects listened to three different musical excerpts, rated to be high in "groove", under varying conditions of movement restriction. These conditions were full restraint (no movement), partial restraint (finger tapping), and no restraint (free movement). Subjects were videotaped throughout the study to ensure compliance with movement restrictions. One-way ANOVA, paired samples t-tests, and regression analyses were performed. The central finding was that the free movement and finger tapping conditions resulted in greater positively-valenced emotional responses (scored on a 9-figure Self-Assessment Manikin) upon music listening than did the still condition. It was also found that the free movement and finger tapping conditions elicited higher ratings of liking (scored on a 5-point Likert scale) of the musical excerpts than the still condition. Findings did not support a relationship between movement restriction and physiological arousal thought to occur with emotional response. It is concluded that movement to a piece of music affects emotional responses upon music listening and that people enjoy pieces of music more when they are able to engage motorically with the music. These findings are relevant to future neuroimaging research into music and emotion as the inability to move freely during neuroimaging procedures may have an impact on emotional responses to pieces of music as well as neural activity involved in these responses. In addition, the role of shifts in attention when one is required to restrict one's movement to music is explored as a possible explanation for decreases in both liking of and positive emotional response to pieces of music.

Auditory-motor Interactions in the Music Production of Adults with Cochlear Implants

Tara Vongpaisal[1] and Daniela Caruso[1]
[1]Department of Psychology, MacEwan University, Edmonton, Alberta, Canada

Cochlear implants (CIs) constrain the perception of pitch in users, yet many continue to participate in musical activities after receiving their implants. To date, there has been no research examining the basis upon which adult CI users integrate electrical hearing input with motor skill to perform a musical instrument. To understand the mechanisms by which CI users learn to produce music, and to what extent auditory and motor systems contribute to this process, we trained adult CI users (n = 5) and hearing controls (n = 15) to learn short sequences on a piano keyboard. We examined how their performance is affected by changes in the melodic and motor sequences in the learned patterns. We adapted a training paradigm used previously to examine the motor and melodic contributions in music production of trained pianists. Using an animated learning procedure that removes the constraints of reading standard music notation, we trained participants to perform short melodic sequences on the piano. We then examined the transfer of learning by measuring the participants' time to perform novel sequences involving a change in the following: 1) melodic pattern only, 2) finger pattern only, 3) melodic and finger patterns, 4) no change. Hearing controls demonstrated the same pattern of difficulty in the transfer of learning across conditions as trained pianists in previous reports. That is, retaining the same melodic and finger patterns in novel sequences enabled the greatest transfer of learning, while those that altered both of these components yielded the poorest transfer of learning. Sequences involving a change in either the melodic or finger pattern yielded intermediate transfer performance. In contrast to hearing controls, there was no systematic pattern in the latency performance of CI participants across transfer conditions. While CI users were able to learn piano melodies successfully under the current training paradigm, our preliminary findings indicate considerable individual differences in the transfer of learning to novel sequences. These differences likely reflect the influence of CI users' unique hearing histories and adaptive strategies in the integration of electrical hearing input and motor skill for music production.

Bobbing, Weaving, and Gyrating: Musician's Sway Reflects Musical Structure

Alexander P. Demos[1], Roger Chaffin[2] & Topher Logan[2]
[1]Department Psychology, University of Illinois at Chicago, Chicago, IL, USA
[2]Department Psychology, University of Connecticut, Storrs, CT, USA

Despite the belief that musicians' bobs, weaves, and gyrations are somehow related to the musical structure, the nature of the relationship has been difficult to assess. Progress has been impeded by four problems: (1) the assumption that musical structure is constant across performances; (2) the complexity of the movements; (3) and the inability of traditional statistical tests to accurately model the multilevel temporal hierarchies involved, (4) and to assess significance. We addressed these problems in a study of the postural sway of two trombonists as each recorded two performances of each of two solo pieces in each of three different expressive styles (normal, expressive, non-expressive). First, we assessed musical structure immediately after each performance by asking the musicians to report their phrasing, which changed across expressive styles and musicians. Second, we measured the complex patterns of movement using recurrence quantification analysis to assess their recurrence (self-similarity), stability and predictability. Third, we used mixed linear models to statistically assess the reliability of recurrence, stability and predictability while taking into account the nesting of phrases within pieces within performances across expressive styles and musicians. Fourth, we also assessed significance of similarity between performance using surrogate techniques to bootstrap confidence intervals by generating surrogate time-series that preserved the autoregressive structure and frequency distribution of the original data. Postural sway was shaped by each musician's phrasing of the music, producing serial position curves that were less pronounced in non-expressive performances and in shorter phrases. Recurrence of sway rose steeply at the beginning of a phrase, peaking before the middle and tailing off gradually. Stability of sway followed the opposite course, falling to a minimum near the middle, before rising and falling again to the end of the phrase.

Short latency effects of auditory frequency change on human motor behavior

Amos Boasson[1], Roni Granot[1]
[1]Musicology department, The Hebrew University, Jerusalem, Israel

Background
A prime element in music, frequency change (FC), being ecologically crucial, may affect within a short latency human motor behavior. In animals, a 'pitch code', used intra and inter species, by senders and receivers, optimizes action, since perceiving direction, magnitude and rate of frequency change enables recognition of identity, intention and urgency. Nevertheless, human motion yielded by FC has been rarely researched.

We hypothesized that FC information affects behavior; that 'Up' or 'Down' may yield differential responses; and that those may be mediated by an 'express' ear to muscle route.

In a previous study, our subjects finger-tapped to isochronous beeps paced at 2Hz, sequenced to present FCs. Tap timing (negative mean asynchrony, NMA) showed an asymmetry: pitch rise yielded earlier taps.

Aims and method
The current study's aim was to examine the immediacy of response to FC. 23 musicians tapped to isochronous beeps. Finger acceleration and EMG activity from antagonist muscles were recorded. Beep pace was doubled (4Hz). FC sequences were intermixed with ones of intensity change to verify that response is not to auditory change per se. Repeated measures ANOVAs were run on the data.

Results
Different NMA behavior in intensity vs frequency conditions implied that the responses are not to auditory change per se. NMA grew (tap occured earlier) in response to rising frequency already on the Tap 1 following change, i.e. within less than 250ms. Further, finger acceleration data revealed, that differential behavior began yet earlier, ca 170ms post change. Aligning data to sound or to tap showed on Tap 1 early triggering of action, but unchanged dynamic profile. The response latency, though short, does not unequivocally show a subcortical 'express' route.

Conclusion
Our results suggest that an ongoing covert human motor response is yielded by frequency change. Such activity may constitute a part of our music perception process, indeed part of music's meaning.

The Wave Transformation: A method for analyzing spontaneous coordination from discretely generated data

Benjamin G. Schultz[1], Alexander P. Demos[2], & Michael Schutz[3]
[1]Faculty of Psychology & Neuroscience, Basic & Applied Neuro Dynamics Laboratory, Maastricht University, Department of Neuropsychology & Psychopharmacology, Maastricht University, The Netherlands
[2]Department of Psychology, University of Illinois at Chicago, USA
[3]Department of Psychology, MAPLE Laboratory, McMaster University, Canada

Coordination between co-performers who produce discrete onsets (e.g., tapping, drumming, piano) has often been measured by performing cross-correlations of the inter-onset intervals (IOI) or through conversion to discrete relative phase and the use of circular statistics (e.g., the resultant vector). The cross-correlation of IOIs can produce coefficients that are relatively low, even when experts perform music together with the goal of synchronizing. Both methods also require several assumptions to be made about the period and/or phase relationships and can be perturbed by missing data. Such assumptions become untenable when people are not intentionally trying to coordinate, but spontaneously align their actions resulting in intermittent coordination. To remedy these issues we have developed the Wave Transformation (WT) Toolbox, a MATLAB toolbox for measuring coordination in discrete onset data. The WT converts the discrete signal into continuous representation similar in appearance to a sinewave, but the peaks of the waves represent the discrete onsets times. The WT signal can be cross-correlated or undergo other more complex signal processing and allows missing data points and unequal numbers of data points. We present simulations where "participants" synchronize with a metronome or with each other to varying degrees. We also provide a method to estimate chance levels of synchrony between a person and a metronome as well as between co-performers. Through direct comparisons of discrete data and the transformed data analyses, we highlight the hazards that can occur when making assumptions for the discrete data. Finally, we demonstrate several applications of the WT for human datasets, including analyses of intentional desynchronization that occurs in Steve Reich's *Drumming*. Our analyses highlight differences in the outcomes of the methods; we discuss which methods are more appropriate under certain conditions.

Hearing the Gesture: Perception of Body Actions in Romantic Piano Performances

Catherine Massie-Laberge,*[1] Isabelle Cossette,*[2] Marcelo M. Wanderley*[3]

*Music Research Department, Schulich School of Music, CIRMMT, McGill University, Canada

[1]catherine.massie-laberge@mail.mcgill.ca, [2]isabelle.cossette1@mcgill.ca, [3]marcelo.wanderley@mcgill.ca

ABSTRACT

Studies conducted on the influence of expressive conditions on the perception of auditors have shown that auditors are better at identifying performers' expressive intentions with visual information than with audio-visual, or audio only. Moreover, velocity of head motion showed meaningful differences between playing conditions (normal, deadpan, exaggerated, immobile). The aims of this work were to assess: whether auditors rely more on motion or auditory cues, in complex Romantic pieces with different expressions, when provided with one perceptual mode at a time; which kinematic parameters convey expressive information; how body part movements influence perception. Eleven pianists performed an excerpt of a Romantic piece in four expressive conditions and were video and audio recorded. Thirty auditors, assigned to a modality (visual, auditory, or both), associated the excerpts to the playing conditions and, using a five-point Likert scale, rated the importance of motion cues (quantity of movement, regularity, velocity) and auditory cues (timing, articulation, dynamic, phrasing) in the association task. They indicated which body parts contributed the most to discriminate the expressions. Auditors were able to decipher pianists' expressions by hearing or seeing the excerpts, but were better in the audio-visual mode. Auditors were confused between deadpan and immobile conditions in the visual mode; and the same occurred for the normal and immobile conditions in the audio mode. Sound dynamics and quantity of motion were the most important cues. Arm and hip movements received respectively the greatest and lowest attention. Auditors found difficult distinguishing immobile from normal playing in the audio mode, which suggests that expressive playing may occur with restricted movements. These findings might equip students with precise informative feedback as regards expression, and help them develop an accurate awareness of auditors' perceptions of body movements.

I. INTRODUCTION

Body movements in the context of musical performances were shown to help auditors extract information about performers' musical intentions and the musical structure (Dahl & Friberg, 2007). But the reasons why it is possible to distinguish performers' personal style are little understood, as well as which kinematic features convey expressive information to auditors.

Two main types of gestures have been discussed in the context of musical performance: effective or instrumental gestures, and sound-accompanying or ancillary gestures (e.g., Cadoz & Wanderley, 2000; Delalande, 1988; Wanderley, 2001). The former represent the gestures responsible of direct modifications to the quality of the sound, and the latter are related to gestures that affect the expression of musical intentions and structural parameters. Jensenius and colleagues (2010) and Godøy (2010) have considered gestures in terms of coarticulated actions in which several gestural actions that follow each other seem to be connected, and therefore may be perceived as one single coherent gesture. For certain instruments, it is not easy to discriminate between strictly sound-producing gestures and expressive, ancillary gestures (sound-shaping gestures). The energy provided by one effector (e.g., the hands) may affect the other parts of the effector apparatus, and involve motion from the arms, shoulders, torso, and even the head.

Many studies used different expressive conditions (normal, deadpan, exaggerated, immobile) and multimodal design to investigate the psychological influence of musicians' gestures on observers. Performers used a varied set of audio cues (e.g., tempo, articulation, timbre, phrasing) to convey diverse expressions, which were decoded by listeners when discriminating between different performance conditions (Gabrielsson & Juslin, 1996; Juslin, 2000). It has also been shown that people can discern expressive intentions through performers' body movements alone. Johansson (1973) was the first to use the point-light method (small light-emitting bulbs fixed to certain parts of the body) to reveal aspects of the perception of body motions. Even when displaying simplified body representation, observers were able to identify actions and intentions with lights, during gait (Montepare, Goldstein, & Clausen, 1987) and knocking movements (Pollick, 2004). Using the technique of point-light displays, Davidson (1994) found that auditors were better at identifying pianists' expressive intentions with visual information than with vision and audio together, or audio only. Velocity of head movement showed meaningful differences between performance conditions (Camurri, Mazzarino, Ricchetti, Timmers, & Volpe, 2004; Davidson, 2007), and between emotions (Castellano, Villalba, & Camurri, 2007; Dahl & Friberg, 2007). In their study on clarinet playing, Nussek and Wanderley (2009) showed that, while performers reduced motion activity in technically demanding musical passages, eliminating body movements (e.g., removing movements from torso or arms and hands) resulted in nearly no effects on observers' judgments of tension, intensity, and fluency of movements. Additionally, clarinetists were shown to play the same piece with idiosyncratic movement velocities. There were still consistencies in their body motions, but the quantity of movement differed between individual body parts.

The interaction between auditory and visual information is also known to impact cognitive and emotional judgments. Looking at piano performances, Camurri and colleagues (2004) found that tempo conveyed phrase boundaries, whereas dynamics were associated to the intensity of the emotion. Phrasing was also related to pattern of tension and release. Furthermore, the timing of motion cues (e.g., the attacks and releases at the beginning or at the end of a phrase) revealed that great amplitude in movements may help observers understand better the performer's expression (Dahl, 2004). Different levels of smoothness of movements may convey different musical intentions and be associated to structural parameters. For instance, a performance with smoother movements is often associated with a sad expression, while jerkiness is related to anger and fear. Smoothness can also be related to the phrasing of music and provide observers with time points of segmentation, as well as a sense of phrasing contour over time (Wanderley, Vines, Middleton, McKay, & Hatch, 2005). While studying the effect of the interactions between visual and auditory information on the judgment of tension and phrasing in clarinet performances, Vines and colleagues (2006) found that high tension was associated to fast, intense, and abrupt movements, whereas low tension was related to smooth and controlled motion patterns. Moreover, certain gestures may help anticipate the beginning of a new section, for instance when the performer initiates a breath while other movements may leave the impression that the phrase extends beyond the end of the note.

Although auditors' perception of expressive parameters conveyed through reduced body representations is impressive, it is unsure which kinematic features allow listeners to perceive this and what specific body movements may influence their perception. In the present study, we seek to understand what influence a piano performance with restrained gestures but usual expressive acoustic characteristics may have on auditors' perception of that performance. The objectives of this work were to evaluate: 1) whether auditors are better in recognizing different performance conditions in complex Romantic piano pieces when provided with one perceptual mode at a time (visual or auditory) or both; 2) how motion cues (i.e., quantity, regularity, velocity) or auditory cues (i.e., timing, articulation, dynamic, phrasing) are efficient in conveying different expressions; and 3) which body parts contribute the most to the discrimination of the different performance conditions.

II. EXPERIMENTAL METHOD

A. Measuring pianists' motion data

Eleven graduate pianist-students (4 females) performed a thirty second excerpt from a well-learned piece of their Romantic repertoire in four different conditions: normal, deadpan, exaggerated, and immobile. Normal performances were played as naturally as possible, and exaggerated performances with an exaggerated level of expression. Deadpan referred to a performance with a reduced level of expression, whereas immobile referred to playing with only the essential movements required to play in a usual manner. Participants could choose the tempo they thought was appropriate to convey the expressive conditions.

Each pianist played one of the following excerpts: Chopin 4th Ballade; Brahms 3rd Piano Sonata; Chopin scherzo no. 2; Schumann Fantasy; Liszt 2nd Ballade; Medtner Sonata Reminiscenza op. 38; Liszt Spanish Rhapsody; Chopin Waltz Op. 64 No. 2; Schubert Impromptu no. 1; Brahms Capriccio op. 76 no. 1; and Chopin Impromptu. Each performance condition was repeated three times (total of 12 performances per pianist). Performances were video recorded with a Sony Wide Angle video camera and audio recorded with a Sennheiser MKH microphone.

B. Auditory and motion cues

Thirty auditors (20 females), with at least 5 years of musical training, were randomly assigned to a modality (vision, audio, or both). They could listen to the excerpts as many times as they wanted in order to associate the musical excerpts to the expressive conditions. After each trial, auditors selected a confidence level on a five-point Likert scale for their answer. They also rated on a five-point Likert scale the importance of motion cues (quantity, regularity, velocity) and auditory cues (timing, articulation, dynamic, phrasing) in discriminating the conditions. Quantity of motion was described as the amplitude of movement, regularity was related to the periodicity of movement and the rhythmic structure, and finally velocity was characterized as the speed and irregularity of the motion (any directional changes). Participants indicated which body parts (i.e., head, torso, shoulders, arms, hands, hips) contributed the most to help them discriminate the expression. Auditors filled out a demographic questionnaire at the end of the experiment.

III. RESULTS

A. Discrimination of expressive conditions

In order to examine the effects of the different expressive conditions, the modalities, and individual pianists on the auditors' ability to discriminate the excerpts, an analysis of variance (ANOVA) with repeated measures was conducted for each of the dependent variables, with the gender and the level of instruction treated as the random variables, and the pianist, modality, and condition treated as the fixed effects.

There is a significant main effect for the conditions ($p < .01$) on auditors' answers, but not for the modalities. Additionally, there is a significant interaction effect between modalities and conditions ($p < .01$); between the modalities and the pianists ($p < .01$); and between the conditions and the pianists ($p < .01$). Figure 1 compares the results obtained for the modalities in each condition. Overall, the expressive condition are better identified in the audio-visual modality (average: 76.1% of good answers), followed by the visual-only modality (72.5%), and finally the audio-only modality (65.5%). The exaggerated performances are fairly well recognized in every modality (median: 90%). The identification of the normal condition in the audio-visual mode varies considerably from 20 % to 100% of good answers and is better recognized in the visual-only mode (median: 80%). However, answers vary between auditors for the audio-visual mode (standard deviation: 11.7) and the audio-only mode (standard deviation: 11.4), but less for the visual-only mode (standard deviation: 5.8). Auditors are often confused between the deadpan and immobile conditions in the visual-only mode and the same occurred for the normal and immobile conditions in the audio-only mode. The auditors' level of confidence in identifying the conditions is higher in the audio-visual modality (weighted average: 78.9%) than in the visual-only mode (weighted average: 73.5%) and in the audio-only mode (weighted average: 66.8%).

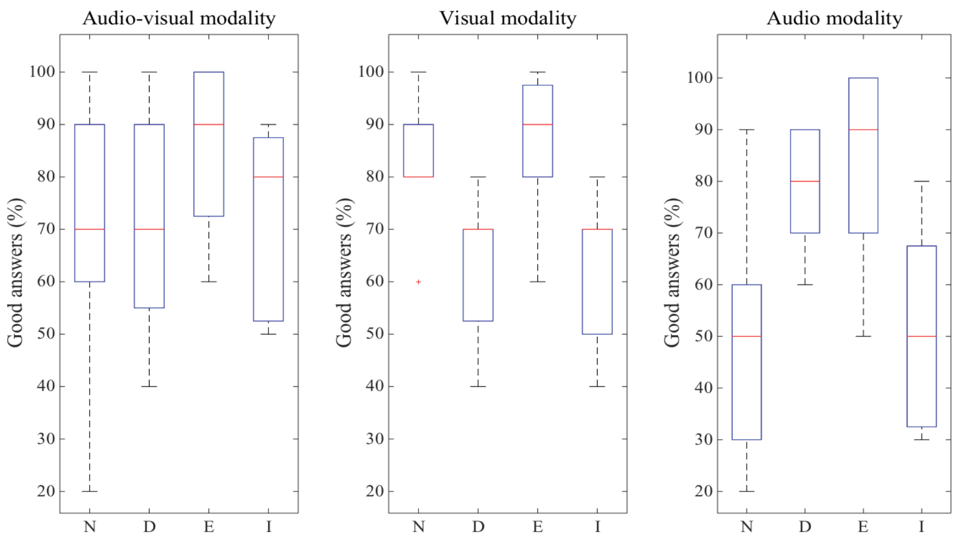


Figure 1. Distribution of the median values per modality illustrating auditors' good answers (in %) with respect to the conditions (N: normal; D: deadpan; E: exaggerated; I: immobile)

Figure 2 shows that the auditors' identification of the expressive conditions varies greatly with the modality when different pianists play. For instance, the conditions are better discriminated in the audio modality than in the audio-visual mode for pianists 4, 8, and 9. However, they are better recognized in the audio-visual mode, followed by the visual-only and the audio-only modes for pianists 2, 3, 5, and 11. Overall, auditors were better at discriminating the condition when pianists 3 and 8 were playing, with respectively an average of 75% and 80% of total good answers, and the lowest average score during pianist 9 playing (63.3% of good answers). Interestingly, auditors deciphered well the conditions in the audio-only modality for pianist 8 (90.0% of recognition).

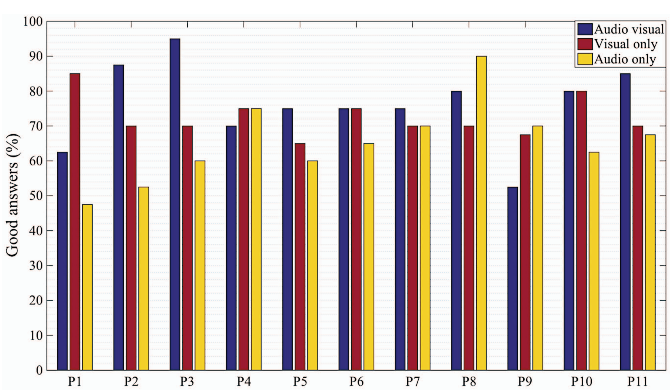


Figure 2. Percentage of auditors' good responses per pianist for each modality

B. Importance of auditory and motion cues

Figure 3 represents the frequency distribution of the auditors' answers (five-point Likert scale) on the importance of audio and motion cues for the identification of the expressive conditions. The vertical line divides the "moderately important" responses in half. Sound dynamic (52.27% very important) and quantity of motion (40.91% very important) are respectively the most important auditory and motion cues. Regularity is perceived as less essential in discriminating the expressive conditions although 19.55% of the responses were rated as "very important". Articulation and timing obtain a very similar response distribution with respectively 28.18% and 24.55% of the answers rated as "very important".

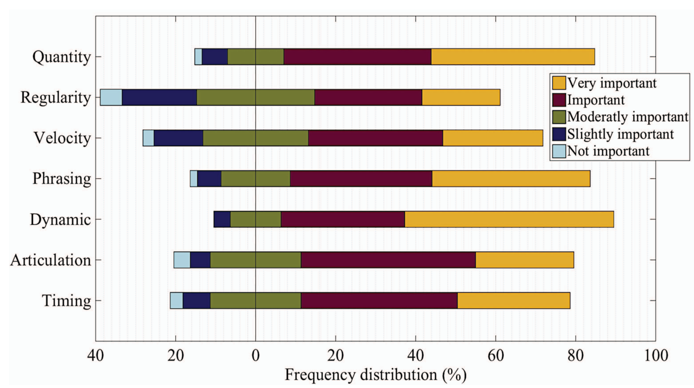


Figure 3. Frequency distribution (in %) of auditors' answers on the importance of auditory and motion cues for condition identification

C. Body parts

In order to identify which body parts are more frequently associated to the motion cues, we measured the frequency distribution (in %) of the auditors' choice of body parts for all pianists together (Figure 4). Arms (71.36% of the time) are the most often body part chosen by auditors to help them discriminate between the conditions, followed by the head (60% of the time), the torso (56.36% of the time), the hands (43.18% of the time), the shoulders (36.36% of the time), and finally by the hips, which are the least often chosen body part (8.64% of the time).

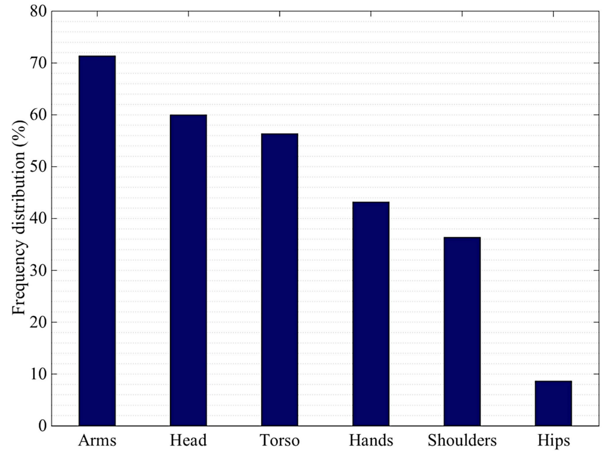


Figure 4. Frequency distribution of auditors' choice of body parts associated to motion cues for all pianists

IV. DISCUSSION

Auditors were able to differentiate pianists' different expressive performances in the three modalities, but were better in the audio-visual modality, which diverges from Davidson's results. Better results were obtained with pianists 3 and 8. These pianists' performances may have played with more contrasts and variations in the sound parameters (e.g., dynamics, voicing, phrasing, articulation, rhythm, timing) and in the motion cues. They may have used greater movement amplitudes for the exaggerated condition, as compared to the normal and immobile conditions.

In the audio-only modality, normal and immobile conditions were not easily discriminated from one another as shown by the lower level of auditors' good answer results, meaning that pianists played in a normal manner even while restricting their movements. The deadpan and immobile performances were often perceived similarly in the visual-only mode. This might be explained by the fact that, even though no instruction was specifically given to pianists as regards the level of movement in that particular condition, the deadpan performances still seem to present a reduced level of motion. This signifies that playing in a deadpan manner may naturally restrict the movements, which may, in turn, suggest that movements are intrinsically connected to expression of pianists. Similar results were also found for normal and immobile performances in the audio-only modality, meaning that pianists played with a normal expression while restricting their movements. This supports Nusseck and Wanderley (2009)'s findings, which pointed out that removing certain kinematic features from the performance might not entail changes in auditors' audio perception of acoustic parameters. Additionally, the responses for the normal condition in the audio-visual mode varied considerably from 20 % to 100% of good answers, whereas in the visual mode, this discrepancy did not occur. This suggests that the removal of auditory information might help discriminate certain conditions.

Auditors' results varied considerably from one pianist to another, which may indicate that one performer may use more amplitude of motion from the arms, whereas another plays with greater changes in velocity patterns from the head. It may be interesting to investigate further pianists' individual differences regarding the use of motion cues to communicate expressive intentions.

The quantity of motion was an essential gestural feature for the auditors to discriminate pianists' expressions. Quantity was rated as the most important cue, followed by velocity, and then by regularity of motion. Results also showed that quantity of motion was strongly connected to the dynamic level of the performance. The broader the motion, the higher the sound level tends to be. Regularity of movement was perceived as less essential, and was often related to the rhythmic structure of the piece, as well as to the articulation.

Davidson (1994) observed that the pianists' hand and wrist movements provided little information about the expressiveness of the performer. Conversely, the auditors of the present study relied strongly on the movements of the arms to help them distinguish between the conditions. Arms were the most often chosen body part, which auditors associated to the motion cues (quantity, regularity, velocity), while in Davidson (1994)'s and Castellano et al. (2007)'s studies the head and torso were. Hips were less frequently chosen than any other body parts. However, this does not mean that the swaying motion from the hips does not yield significant differences between the conditions. Although motion from the hip region may not be as visually perceptible as the head motion, it still constitutes the pivotal point for all upper torso and head movements. In fact, it might be an essential expressive parameter in piano performance that conveys information about the musical structure. This still needs to be investigated in further studies.

V. CONCLUSION

The present study demonstrates that:

- auditors are able to decipher pianists' different expressive performances in three modalities (audio-only, visual-only or audio-visual) but are better in the audio-visual modality;
- pianists can play expressively while restricting their movements;
- dynamic and quantity of motion are important parameters of expression and;
- arm motions provide meaningful information to discriminate between different expressions.

Further work is needed to understand better the relationships between pianists' expressive communication in Romantic piano repertoire, auditors' understanding of expressive parameters, and the structural features of the music.

The analysis of bodily movement through performance conditions can provide more information as regard the functions of gestures in musical performance. This information could benefit the field of piano pedagogy, and also serve to design new technologies to help student-performers become aware of the influence of motion cues on communication of various expressive intentions.

ACKNOWLEDGMENT

We gratefully acknowledge financial support from the Social Sciences and Humanities Research Council of Canada. We would also like to thank all the performers and auditors for their participation and collaboration.

REFERENCES

Cadoz, C., & Wanderley, M. M. (2000). Gesture-music. In M. M. Wanderley & M. Battier (Eds.), *Trends in Gestural Control of Music* (pp. 71-94). Paris: IRCAM - Centre Pompidou.

Camurri, A., Mazzarino, B., Ricchetti, M., Timmers, R., & Volpe, G. (2003). Multimodal analysis of expressive gesture in music and dance performances. In A. Camurri & G. Volpe (Eds.), *Gesture-based Communication in Human-Computer Interaction* (pp. 20-39). Berlin: Springer Verlag.

Castellano, G., Villalba, S. D., & Camurri, A. (2007). Recognising human emotions from body movement and gesture dynamics. In A. C. R. Paiva, R. Prada & R. W. Picard (Eds.) *Affective Computing and Intelligent Interaction* (pp. 71-82). Berlin: Springer Verlag.

Dahl, S. (2004). Playing the accent-comparing striking velocity and timing in an ostinato rhythm performed by four drummers. *Acta Acustica united with Acustica*, *90*(4), 762-776.

Dahl, S., & Friberg, A. (2007). Visual perception of expressiveness in musicians' body movements. *Music Perception: An Interdisciplinary Journal*, *24*(5), 433-454.

Davidson, J. W. (1994). What type of information is conveyed in the body movements of solo musician performers. *Journal of Human Movement Studies*, *6*, 279-301.

Davidson, J. W. (2007). Qualitative insights into the use of expressive body movement in solo piano performance: a case study approach. *Psychology of Music*, *35*(3), 381-401.

Delalande, F. (1988). La gestique de Gould: éléments pour une sémiologie du geste musical. In G. Guertin (Ed.), *Glenn Gould, Pluriel* (pp. 83-111). Montréal: Louise Courteau éditrice.

Gabrielsson, A., & Juslin, P. N. (1996). Emotional expression in music performance: Between the performer's intention and the listener's experience. *Psychology of Music*, *24*, 68-91.

Godøy, R. I., Jensenius, A. R., & Nymoen, K. (2010). Chunking in music by coarticulation. *Acta Acustica united with Acustica*, *96*(4), 690-700.

Jensenius, A. R., Wanderley, M. M., Godøy, R. I., & Leman, M. (2010). Musical gesture: Concepts and methods in research. In R. Inge Godøy & M. Leman (Eds.), *Musical Gestures: Sound, Movement and Meaning* (pp.). New York, NY: Routledge.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics*, *14*, 201-211.

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1797-1813.

Montepare, J. M., Goldstein, S. B., & Clausen, A. (1987). The identification of emotions from gait information. *Journal of Nonverbal Behavior*, *11*, 33-42.

Nusseck, M., & Wanderley, M. M. (2009). Music and motion—How music-related ancillary body movements contribute to the experience of music. *Music Perception: An Interdisciplinary Journal*, *26*(4), 335-353.

Pollick, F. E. (2004). The features people use to recognize human movement style. In A.Camurri & G.Volpe (Eds.), *Gesture-Based Communication in Human-Computer Interaction* (pp.10-19). Berlin: Springer Verlag.

Vines, B. W., Wanderley, M. M., Krumhansl, C. L., Nuzzo, R. L., & Levitin, D. J. (2003). Performance gestures of musicians: What structural and emotional information do they convey? In A. Camurri & G. Volpe (Eds.), *Gesture-Based Communication in Human-Computer Interaction* (pp. 468-478). Berlin: Springer Verlag.

Wanderley, M. M. (2001). Quantitative analysis of non-obvious performer gestures. In I. Wachsmuth & T. Sowa (eds.), *Gesture and Sign Language in Human-Computer Interaction* (pp. 241-253). Berlin: Springer Verlag.

Wanderley, M. M., Vines, B. W., Middleton, N., McKay, C., & Hatch, W. (2005). The musical significance of clarinetists' ancillary gestures: An exploration of the field. *Journal of New Music Research*, *34*(1), 97-113.

Participant Attitudes toward Dalcroze Eurhythmics as a Community Fall Prevention Program in Older Adults

Hyun Gu Kang[1], Rodney Beaulieu[2], Shoko Hino[1], Lisa Brandt[1] and Julius Saidro[1]
[1]Kinesiology, California State University San Marcos, San Marcos, California, USA
[2]Human Development, California State University San Marcos, San Marcos, California, USA

Dalcroze Eurhythmics (DE) is a music education program that that incorporates cognitive-motor skills cued by music. Participation in DE improved gait rhythm, cognitive ability, and mood, and was very effective in reducing the risk of falls in older women in Geneva. Our goal was to determine the perception of DE as a fall prevention and brain fitness program in order to improve its attraction and efficacy in an American population. A series of 10 weekly one-hour classes in DE was offered at the local senior center by a licensed Dalcroze instructor. Classes consisted of activities and games cued by music to challenge attention, memory, coordination, balance, social interaction, and coordination. Participants were queried using a focus group regarding the class. 24 older adults (19 Females) of mean age 76 enrolled in the program. 13 participated in the focus group. Strengths of the program noted were that the "wide variety of exercise and games which included the mind and the body", "listening to music … because we have to keep pace and slow down," the presence of "young people that helped with the class, " "music from our era which was fun," and the activities that "keep us from sitting-make us move." Many noted that the interactive aspect of the class as a strength, and is not intimidating like other fitness programs. One weaknesses of the program was lack of progression of activities for high functioning participants as noted by half of the participants. Another was that the room felt crowded on days with high attendance. To improve the program, future offerings will involve more students, progressive level classes, and smaller classes or larger spaces.

Evaluating the Audience's Perception of Real-time Gestural Control and Mapping Mechanisms in Electroacoustic Vocal Performance

J. Cecilia Wu[1] and Matt Wright[2]
[1]Media Arts and Technology Department, University of California, Santa Barbara, CA, USA
[2]Music Department, Stanford University, Stanford, CA, USA

This paper presents the first empirical evaluation of a digital music instrument (DMI) that is designed for augmenting electroacoustic vocal performance from the audience's perception. We strive to understand the audience preference for the DMI, a Tibetan Singing Prayer Wheel, which simultaneously processes vocals based on the performer's hand gestures; and it's mapping between vocal processing parameters and sensed horizontal spinning gesture, in particular the time-varying rotational speed. We hypothesize that the audience's level of perceived expression and their engagement level of a performance increase when: 1. Using synchronized gestures mapping to control the vocal processing; 2. Using what we consider intuitive gesture mapping to control the vocal processing. In both experiments, we made two versions of each song's audio and inserted them as alternative soundtracks for a video recording of a single performance. This methodology eliminates possible confounds associated with comparing different performances; in our case every comparison is between identical videos of the same performance but with alternative gesture to effects-control mappings. Experiment 1 compared the original mapping, our "synchronicity" condition, against a desynchronized alternative, with control functions taken from an unrelated performance. Experiment 2 compared the original mapping (slow rotation producing a more natural sound and faster rotation causing a progressively more intense granular stuttering effect); versus the opposite. We then presented all six songs to two groups of participants, and evaluated their responses via questionnaire. In both experiments, viewers reported overall higher engagement, preference and perceived expressiveness for the original versions, and prefer our original design. Through evaluating these control and mapping mechanisms, we aim to push the discussion towards developing gesture theory in electroacoustic vocal performance.

Walking to music: How instructions to synchronize alter gait in good and poor beat perceivers

Lucy M. McGarry[1], Emily A. Ready[3], Cricia Rinchon[1,2], Jeff D. Holmes[3], Jessica A. Grahn[1,4]
[1]Brain and Mind Institute, University of Western Ontario, Canada
[2]Schulich School of Medicine and Dentistry, University of Western Ontario, Canada
[3]Health and Rehabilitation Sciences, University of Western Ontario, Canada
[4]Department of Psychology; University of Western Ontario, Canada

Background
Music-based rhythmic auditory stimulation (RAS) is a cueing technique used in gait rehabilitation, typically involving synchronizing footsteps with the beat. However, instructions to synchronize may be detrimental to individuals who have difficulty perceiving the beat in music, such as people with Parkinson's disease, by causing an increase in cognitive load.

Aims
Here we evaluate how beat perception ability influences gait responses to RAS when instructed to synchronize steps with the beat versus when permitted to walk freely. We hypothesized that walking freely to music without instructions to synchronize may benefit poor beat perceivers by reducing dual-task load.

Methods
84 healthy adult participants walked on a pressure-sensor walkway to music that they rated to be high in groove (the 'desire to move to the music'). Beat perception ability was evaluated using the Beat Alignment Test from the Goldsmiths Music Sophistication Index v1.0.

Results
As predicted, instructions to synchronize facilitated gait in good beat perceivers, whereas walking freely facilitated gait in poor beat perceivers. This instruction x beat perception interaction was statistically significant for stride width, an indicator of walking confidence and balance, and approached significance for stride length, an area of particular concern in Parkinson's. Post-hoc tests on song rating data indicate that gait improvements in good beat perceivers were linked to perceptions of high beat salience and groove in the music, whereas improvements for poor beat perceivers were linked more closely to enjoyment.

Conclusions
Results support the premise that RAS should be tailored to beat perception ability: good beat perceivers benefitted more from synchronizing, whereas poor beat perceivers benefitted more from freely walking to the music.

A comparison of three interaction measures that examine dyadic entrainment

Marc R. Thompson[1], Birgitta Burger[1], and Petri Toiviainen[1]
[1]Department of Music, University of Jyväskylä, Finland

The body movements involved in dyadic music performance usually fall into two broad categories; sound-producing movements, which are planned and highly coordinated, and ancillary gestures, which are improvised yet essential for communicating musical intentions. Previous work has shown that these types of movements are correlated to some extent as they both reflect a performance's attributes such as musical structure and tempo. However the categories differ enough that they may require different interaction analysis methods when assessing interpersonal coordination. This study investigates which interaction analysis method is best suited for each movement category, and also considers a range of movements from simple coordinated actions to musical performances. In a series of trials, two violinists were motion-captured while performing short coordinated gestures (e.g. mirrored and unmirrored limb extensions), uncoordinated improvised actions (e.g. free gestures and dancing), and a short work for two violins. Interpersonal coordination is evaluated using three different methods that uncover generalized relationships between data sets: Canonical correlation analysis (CCA: identifies linear relationships by calculating the shared variance between data sets), mutual information (MI: identifies mutual dependence of two data sets) and joint multivariate synchrony (JMS: identifies the joint entropy of two data sets). Though some work has started with investigating coordination using CCA, progress on this project is ongoing. Generally, we expect to find that the effectiveness of a given interaction analysis method will be based on the nature of the trial. For instance, because CCA assumes linear relationships between data sets, this technique would be best suited for evaluating the degree of synchrony between coordinated instrumental movements, whereas MI makes no such assumptions, and could therefore be used for assessing interaction in improvised or ancillary gestures.

Eye movements while listening to various types of music

Marek Franěk[1], Denis Šefara[1] and Roman Mlejnek[2]
[1]Faculty of Informatics and Management, University of Hradec Králové, Hradec Králové, Czech Republic
[2]The Prague Conservatoire, Prague, Czech Republic

A large body of research was devoted to the way in which music shapes the perception of the visual stimuli. The recent research demonstrated that the evaluation of outdoor scenes can be affected by characteristics of the listened music. The present study using eye tracking method was intended to investigate differences in perception of urban and natural scenes while listening to two distinctive types of music. The effects of (1) motivational music, which has a fast tempo and strong rhythm and (2) non-motivational music, which is slower and with no strong implication to movement on perception of various urban and natural scenes were investigated. We were interested, whether particular types of music can influence perception of the scenes in terms of a number of eye fixations, durations of fixations, and an extent of perceived visual information. 66 undergraduates aged from 19 to 25 years participated in the study. They viewed 25 photographs successively. Simultaneously, they listened to motivational or non-motivational music. The no-music condition was used for a control. Participants' gaze was recorded using the Tobii X2-60 eye tracker. Durations and a number of fixations in the photographs were analyzed. We did not find any significant differences in the extent of perceived visual information. However, the number of fixations was significantly higher in the no-music condition compared to the music conditions. The durations of fixations were the shortest in the no-music condition. Moreover, the duration of fixations of the participants listening to motivational music were significantly longer relative to those listening to non-motivational music. The data suggested that listening to music while viewing the photographs required larger cognitive load, which resulted in longer durations of eye fixations. The research could enhance our understanding of phenomena associated with music listening while walking or running in an outdoor environment.

Periodic Body Motions as Underlying Pulse in Norwegian Telespringar

Mari Romarheim Haugen[1]
[1]Department of musicology, University of Oslo, Norway

The pulse level in music is often described as a mental reference structure against which we perceive rhythm. The pulse is often defined as a series of isochronous beats, however this conception is challenged by music styles that seem to feature an underlying pulse that consists of beats of uneven durations, such as certain traditional Scandinavian dance music genres. A considerable part of the traditional dance music of Sweden and Norway are in so-called *asymmetrical meter*—that is, the beats in a measure are of uneven duration.

In this study we focus on one specific music style, namely Norwegian telspringar. The intimate relationship between music and motion is often highlighted in rhythm studies of these music styles, so we wanted to investigate whether performers' periodic body motions in telespringar performance could be understood as an expression of the underlying pulse.

We report from a motion capture study where three professional performers, one fiddler and two dancers, participated. The participants' body motions were recorded using an advanced optical infrared motion capture system.

Sound analysis showed cluster of peaks at beat level, which made it difficult to determine a duration pattern based on the sound signal. Motion analysis of the fiddler's integrated foot stamping, on the other hand, showed clear acceleration peaks revealing a very stable *long – medium – short* duration pattern that seemed to be synchronized with the beat related sonic events in the music. Analysis of the dancers' vertical hip motions also revealed a stable motion pattern that seemed to be synchronized with both measure level and beat level indicated by the fiddler's foot stamping.

The results show that underlying asymmetrical structure is indicated by the fiddler's foot stamping and corresponds to the dancers' vertical hip motions. This confirms the view that the underlying pulse in telespringar consists of beats of uneven duration.

Bouncing synchronization to vibrotactile music in hearing and early deaf people

Pauline Tranchant[1], Martha Shiell[2], Marcello Giordano[3], Alexis Nadeau[2], Marcelo Wanderley[3], Isabelle Peretz[1] and Robert Zatorre[2]

[1]International Laboratory for Brain Music and Sound Research, University of Montreal, Montreal, Quebec, Canada

[2]Montreal Neurological Institute, McGill University, Montreal, Quebec, Canada

[3]Input Devices and Music Interaction Lab, Centre for Interdisciplinary Research in Music Media and Technology, McGill University, Montreal, Quebec, Canada

In dance, people synchronize their movements with the musical beat. Since music is primarily experienced through the auditory sense, this form of sensorimotor synchronization has exclusively been studied using auditory stimuli. However, music can also be experienced through the tactile sense, when objects that generate sounds also generate vibrations. One typical example is the vibration of the floor induced by the subwoofers in dance clubs. This is of particular relevance for deaf people, who rely on non-auditory sensory information for beat synchronization. In the present study, we aimed to investigate beat synchronization to vibrotactile electronic dance music in hearing and deaf people. We tested 8 deaf and 15 hearing individuals on their ability to bounce in time with the tempo of vibrotactile stimuli (no sound) delivered through a vibrating platform. The corresponding auditory stimuli (no vibration) were used in an additional condition with the hearing group. We collected movement data using a camera-based Motion Capture system, and subjected it to a phase-locking analysis to assess synchronization quality. The vast majority of participants were able to precisely time their bounces to the vibrations, with no difference in performance between the two groups. In addition, we found higher performance for the auditory compared to the vibrotactile condition in the hearing group. Our results thus show that accurate tactile-motor synchronization in a dance-like context does occur regardless of auditory experience, though auditory-motor synchronization was of superior quality.

Conducting – an Intuitive Gestural Language? Exploring the Actual Impact of Conducting Gestures.

Sarah Platte,[*1] Morten Schuldt-Jensen[#2]

[*] *MIT Media Lab, Massachusetts Institute of Technology, Cambridge, MA, USA*

[#] *Department of Choral and Orchestral Conducting, University of Music, Freiburg, Germany*

[1]platte@media.mit.edu, [2]schuldtjensen@gmail.com

ABSTRACT

Being the only musician without a sounding instrument, the conductor is the target of admiration as well as harsh criticism, and although allegedly of great importance to any ensemble performance, very few people – conductors themselves included – have more than a vague notion of what the person with the baton is actually doing. However, given there is a connection between the conductor's gesture and the musicians' reaction, there must also be explorable mechanisms making this special sign language understandable across instrumental, cultural and language borders. This study aims to identify and map the physical parameters in this subtle – even mythical – communication process.

Based on a detailed literature review and an analysis of internationally approved books on conducting technique we decided to test 3 different archetypes of gestures performed in the same setting by the same conductor to rule out possible differences in "conductors' personality". The first study measured timing and pressure of touch-events on a touch-sensor. Subjects were asked to tap the beat while shown videos of the three different conducting patterns. In the second study we asked violinists to play single notes following the same videos as in study 1, but here we additionally measured differences in sound-quality.

By means of both studies we could prove a consistent and direct correlation between the gestures and muscle-tension of the conductor on one side, and the physically measurable reactions of musicians, onset-precision, duration and sound quality on the other side.

I. INTRODUCTION

Expert or fraud – opinions differ widely when it comes to the profession of the conductor. The powerful person in front of an orchestra or choir attracts both hate and admiration. But what actual influence does a conductor have on the musician's body and the sounding result?

A large proportion of the myth around the conductor is inextricably linked to his individual influence on the sound of an orchestra. In this context, many reports of the conductor's unique influence on the orchestra exist. For example, Arthur Nikisch is said to have "had the uncanny ability to get a ravishingly beautiful sound out of an orchestra without saying a word" (Seaman, 2013). A commonly used explanation of a specific sound of an orchestra is the aura of a charismatic conductor. The most famous example is probably Wilhelm Furtwängler, who was known to change the orchestra's sound just by entering the room, even when someone else was conducting (Lawson, 2003). Recent books on the role of the conductor refer to more scientific concepts, such as mirror neurons or emotional contagion, as an explanation for the apparently inexplicable influence of conductors on the sound of an orchestra. Unfortunately, so far no research exists on conductor-musician communication in these fields. Wolfgang Hattinger (2013) concludes that there are no rationally explicable or controllable energies and transferences in play. But, besides aura, personality and emotional contagion, could gestures themselves also have an influence on the sound?

Aiming to throw light on the fundamental principles of this special gestural language, the presented research approaches the communication between conductor and musician as a matter of physics and as an analogic – rather than a symbolic language. By means of the presented studies, we can prove an influence of different shapes and sizes of conducting gestures on the length and volume of the elicited note and measurable differences in sound quality.

II. BACKGROUND

A. Related Work

Related research on this topic can be found in the field of psychology – multiple studies could prove that humans are capable of integrating different senses multimodally, due to subconscious perception. Michael Haverkamp (2001) discriminates between different levels of cognitive connections within auditive and visual perception, using an either intuitive or cognitive classification process. This capability of integrating multimodal perception has been essential for the evolutionary advantage of the human race.

There seem to be many abstract multimodal analogies between sound and color, form, size or emotion. Two examples of research on this phenomenon are Manfred Clyne's "essentic forms" (1978) and Wolfgang Köhler's "Maluma & Takete" (1929 & 1947). Clynes developed a machine called the sentograph, which measures people's gestural responses to various pieces of music. Based on these findings he claims the existence of distinct expressive forms for a number of emotions. In 1929, Köhler explored a non-arbitrary mapping between speech sounds and the visual shape of objects. By showing participants two forms similar to those in Figure 1 and asking, which shape was called "takete" and which "maluma", respectively, he came to the result, that there is a strong preference to link the rounded shape to "maluma" and the jagged shape to "takete".

Köhler's experiment has been repeated in various settings and variations (Jankovic, 2006, Nielsen & Rendall, 2013, Oakley & Brewster, 2013). An equivalent consistency in the combination of gestural expressions and sounding reaction in case of a conductor-musician communication, however, has not been explored yet.

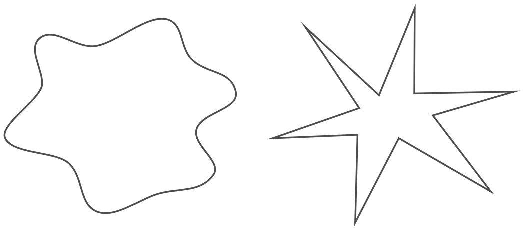


Figure 1. Maluma (left) and Takete (right)

Within the field of conducting, research by Serrano (1993) and Schuldt-Jensen (2016) is particularly relevant for our study design. Serrano's experimental investigation showed a high predictability of conducting patterns with parabolic paths (arcs with ends down) and a lower predictability of inverse parabolic paths (arcs with ends up).

Schuldt-Jensen's analysis indicates the exact opposite: he differentiates between four categories of characteristics of the different conducting patterns:"1) levels of the beating point 2) form of the trajectories 3) form of upper and 4) form of lower vertical turning point". Based on this analysis of general beat patterns, Schuldt-Jensen attributes a high predictability to convex trajectories and a lower predictability to concave conducting patterns.

Furthermore, Schuldt-Jensen hypothesizes an intuitively evoked quasi staccato reaction in case of purely concave conducting gestures, and a quasi legato as reaction to convex trajectories. A similar assumption can also be found in several teaching books on conducting, assigning gestures with sharp, jagged beating points to staccato and smooth, rounded ones to legato (e.g. Green, 1992, Farberman, 1997, Bowen, 2005). In addition, a majority of teaching books also suggest to modify the size of the conducting gesture to evoke different dynamics, using big gestures for forte and small gestures for piano (e.g. Hinton, 2008, Schuldt-Jensen, 2008).

B. Methodology

Based on the literature review, we decided to use three different archetypes of conducting patterns for our experiments: Concave, convex and mixed (see Figure 2). Concave and convex gestures are simplified versions of those used by Serrano (1993), reduced to the pure shape as being convex or concave as defined by Schuldt-Jensen (2016). The mixed gesture is an often taught combination of a concave and a convex trajectory, for example seen in Phillips (1997), Lumley & Springthorpe (1989) and Göstl (2006).

In addition to the three basic patterns, we were also using a big and a small version of the convex pattern to investigate possible differences in volume of the evoked reaction. In relation to onset synchrony the question of predictability becomes especially interesting when it comes to sudden tempo changes. Therefore, as a supplement to measuring the reaction on the three different conducting patterns in constant tempo, we included versions with changing tempo for all three gestural archetypes to see whether response time differs.

Figure 2. Concave (left), convex (middle) and mixed (right) conducting patterns with red beat points.

In order to isolate the visual information of a conducting gesture as best as possible, we decided to use only the conductor's right hand, and not the left hand which is in some conducting schools assigned to exclusively show the expressive content of the conveyed information. Furthermore, we blurred the conductor's face in the video to exclude any influence from facial expressions.

For the second study, we decided to investigate two additional factors. Firstly, the conductor's hand posture: When observing hand postures of conductors, four variations can be found: the palms of the hands face either downward (pronation), sideward (neutral), upward (supination) or both hands are showing different hand postures. In order to investigate potential differences in reaction on pronated, supinated and neutral hand positions, we recorded conducting gestures with all three hand positions, using the same convex form of trajectory and conducted in a constant tempo.

Secondly, when it comes to conducting a fermata, two contradicting approaches can be found in teaching books: The first method suggests to completely stop the movement on the beat where the fermata starts, and then hold the position as long as the conductor wants the note to sound (McElheran, 2004, Farberman, 1999, Galkin, 1988). In contrast, the second approach recommends to continue with a slow movement during the fermata to support a sustained sound (Maiello et al., 1996, Phillips, 1997, Schuldt-Jensen, 2008 and Colson, 2012). In order to explore potential differences in muscle tension and sound quality, we presented participants with both versions of the fermata and asked them to play the fermata "as shown by the conductor".

III. DATA COLLECTION

A. Touch-Sensor Study

The experimental setup of the first study consisted of a screen, showing videos of prerecorded conducting gestures, headphones for audio feedback and a touch-sensor (Force-Sensitive Resistor (FSR)), which detected physical pressure.

The within-subjects experiment with two randomized sets included 19 healthy subjects of both genders (12 male / 7 female) with an age between 19 and 32 ($\bar{x}$ 26.5). 14 participants were musicians/singers who had already or currently played or sung with a conductor; 5 were non-musicians who had never worked with a conductor.

During the experiment the subject was instructed to follow videos of varying conducting gestures by playing a sound elicited by pressing the touch sensor. Every tapping-event induced an audio feedback, the length and force of the subject's pressure mirroring the length and volume of the evoked tone respectively. The on- and offset-time of each tapping event and the amount of pressure on the sensor were recorded. The collected sensor data were synchronized with the recorded data of the shown conducting gestures to allow further investigations.

The participants were asked to follow the tempo of the conductor and to intuitively play – in terms of length and volume – as shown by the conductor. The proceeding tasks can be subdivided into three categories:

1) Concave, Mixed, Convex. Videos of the three basic movement types were shown, conducted in constant tempo. Measurements of length and amount of pressure on the FSR sensor were collected, and mean and standard deviation (STD) were calculated.

2) Dynamics & Articulation. For dynamics, two conducting gestures with identical form, but differing in size, were shown to measure differences in the applied pressure on the touch sensor. As for the effect on the articulation, the two contrasting conducting patterns (concave and convex from category 1) were used, measuring potential differences in the length of pressure on the sensor.

3) Predictability. Three videos of concave, convex and mixed conducting gestures with changing tempo were used to investigate the predictability of the different types of trajectories, indicated by the standard deviation (STD) from the calculated beat point.

B. Violin Study

Violinists represent the largest instrumental group of orchestral musicians. As target group for our second experiment they are especially interesting, because all main body parts involved in the process of playing the violin are visible and accessible for sensors. Moreover, the violin offers a comparatively open sound configuration; timbre varies depending on the bow's speed, pressure, angle, and position of contact, as well as on the left hand's finger pressure on the strings.

The experimental setup for this study consisted of a screen, showing videos of the prerecorded conducting gestures and a Zoom H2 device to record audio in 44100 Hz stereo, 32 bit wave-format. As orchestra musicians usually play seated, we also asked our participants to sit during the experiment. While following a protocol of single tasks, the participants wore three MYO Armbands (Thalmic Labs, 2016), one on the right and left forearm respectively and the third on the right upper-arm.

The within-subjects experiment with two randomized sets included 8 healthy subjects of both genders (1 male / 8 female) with an age between 19 and 63 ($\bar{x}$ 31). All 8 participants are active violinists with at least 8 years of experience of playing under a conductor; 4 participants are professional violinists and members of professional symphony orchestras, 4 play on a semi-professional level.

After mounting the sensor armbands, the subject was instructed to follow videos of varying conducting gestures by playing a d'' with the third finger on the a-string in half notes. The recorded audio and collected sensor data was synchronized with the recorded data of the conductor to allow further investigations.

The proceeding tasks can be subdivided into three categories:

1) Concave, Convex and Mixed. The first category investigates the intuitive reaction to the different forms of conducting gestures in terms of sound quality and muscle tension.

2) Pronation, Supination and Neutral Hand Posture. With different forearm rotations we aim to explore potential influences of hand postures on the sound quality and muscle tension of conductor and musician.

3) Fermata. This category is designed to show possible differences in sound quality in cases of a fermata with respectively without movement.

In both studies, we executed every task twice in a randomized sequence to prevent order effects, and we closed the experiments with a short questionnaire, asking for background information on age, gender, profession, and musical education/experience.

IV. RESULTS

A. Touch-Sensor Study

Measuring intuitive reactions on different gestural archetypes, significant differences occurred in all measured categories.

1) Concave, Mixed, Convex. The intuitively evoked pressure on the touch sensor, which represents the produced volume of the note, differed significantly when comparing concave/convex ($p=0.00007$) and concave/mixed ($p=0.009$) gestures. The difference between convex and mixed is not significantly distinctive ($p=0.4$). The mean pressure is the highest in convex (374.8), followed by mixed with 359.5 and the lowest in concave (337.9). However, the envelope of a pressure event indicates that the difference in mean pressure should rather be described as the overall energy during one specific touch event than just mean pressure itself, as there is a longer plateau on a high pressure level in convex and mixed gestures as compared to the sharp rise and immediate decline in the concave gesture (see Figure 3).

When it comes to the length of the evoked note, differences are considerable. There were statistically highly significant differences between mean values of concave, mixed and convex as determined by a one-way ANOVA test ($p = 0.0000000014$). As is evident from the envelopes in Figure 3, the length of the resulting note of a concave gesture (295.5 ms) is less than half of those of mixed (620.4 ms) and convex (731.8 ms).

2) Dynamics & Articulation. Different forms and sizes of conducting gestures lead to highly significant differences in pressure length and force on the touch-sensor.

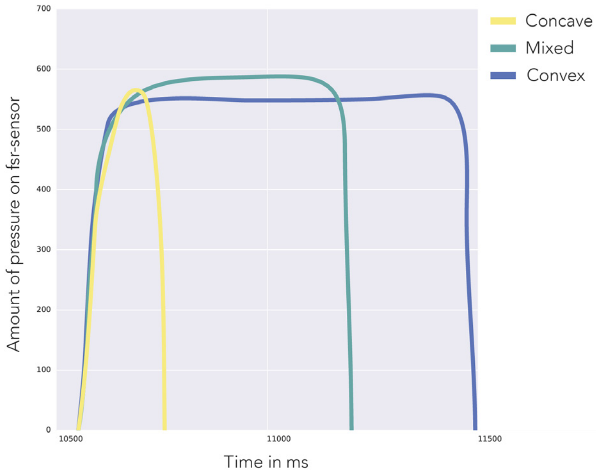
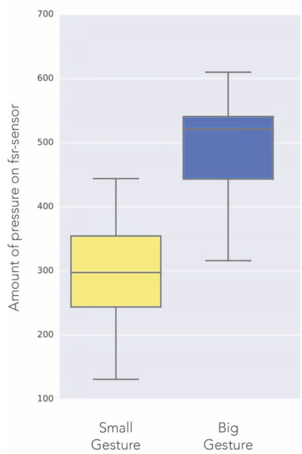


Figure 3. Envelopes of touch events



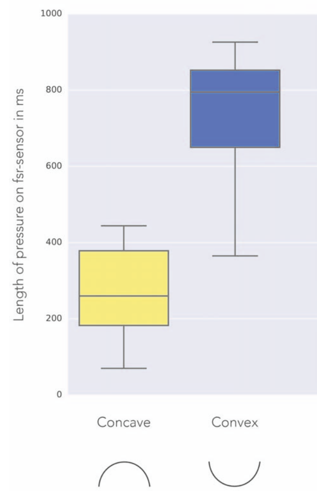


Figure 4. Comparison of intuitively evoked volume.

Figure 5. Comparison of intuitively evoked length of pressure in ms.

As for different sizes of gestures (see Figure 4), the mean pressure proves to be 293.5 mV for the small and 489.5 mV for the big gesture (p = 0.006 -20).

When it comes to envelope form (corresponding to the musical category articulation), the convex gesture evokes a significantly longer pressure duration than the concave gesture ($p = 0.003^{-6}$) (see Figure 5).

3) Predictability. Videos of concave, convex and mixed conducting gestures with changing tempo are used to investigate the predictability of the different types of trajectories. By means of a frame-by-frame video analysis we first extracted the conductor's beat points, which are located at the lower turning point of each trajectory. These beat points are then compared with the participants' onset times of each pressure event.

The observed differences in predictability of the three gestural archetypes can be confirmed in significance by calculating the one-way ANOVA of the delays' standard deviation (p = 0.04).

B. Violin Study

1) Concave, Convex and Mixed. Measuring intuitive reactions on different gestural archetypes, significant differences occurred in the measured intensity of the evoked tone.

By means of the Sonic Visualiser (2016), we calculated the intensity of the signal, i.e. the sum of the magnitude of the FFT bins of 7 sub-bands. As shown in Figure 6, the highest average intensity was induced by a convex gesture (66.4), followed by mixed (56.7) and concave (34.1). The observed differences in intensity of the three gestural archetypes can be confirmed in significance by calculating the one-way ANOVA (p = 0.0001). These results confirm the findings of intuitively evoked volume in form of pressure in the touch-sensor study. Interestingly, an analysis of the conductor's muscle tension in the right forearm shows contrary results: The conductor uses the highest average muscle tension with the concave gesture (90.2), followed by mixed (41.5) and convex (33.7). Although the resulting intensity is consistent trough all participants, measurements of muscle tension of the musicians do not give significant results.

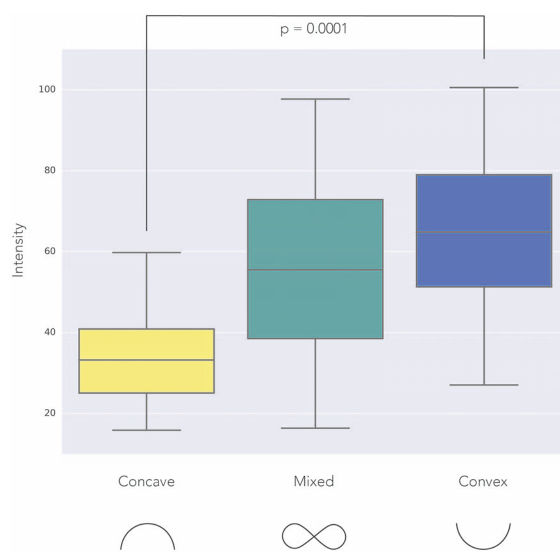


Figure 6. Intensity comparison of evoked tone

2) Pronation, Supination and Neutral Hand Posture. Different forearm rotations, conducted with the same (convex) trajectory, show significant influences on the evoked intensity.

Differences between neutral hand posture and pronation show the highest significance with p=0.0003. Also, the divergence between neutral and supination is significant (p=0.02), while the comparison of the results in cases of supination and pronation is non-significant (p=0.1). As observed in category 1 as well, measurements of the conductor's muscle tension in his right lower arm show contrasting results: as depicted in figure 7, the neutral hand posture shows the lowest muscle tension in the conductor's forearm while apparently evoking the highest intensity in (the violinist's) sounding reaction. At the same time, a high muscle tension in pronation evokes the lowest intensity. As in Category 1, although highly consistent in the resulting intensity, the corresponding measurements of muscle tension of participants do not give significant results.

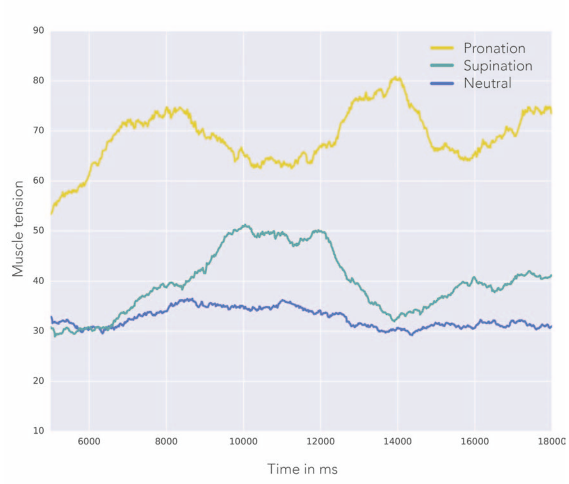


Figure 7. Muscle tension in the conductor's right forearm.

3) Fermata. Measurements of intensity in the two different versions of the fermata show significantly (p=0.01) higher values (mean intensity: 20.8) for the fermata with movement compared to the static version (mean intensity: 17.7). Also, significant differences in mean muscle tension of the conductor indicating a slightly higher value for the fermata with movement.

The movement of the conductor's arm during a fermata seems to not only have an influence on the mean intensity, but the envelopes of intensity over time differ considerably: The version with movement starts with a high intensity that remains relatively stable until the middle of the fermata, before decreasing until the end of the note (see figure 8). Figure 9 shows a typical fermata without movement, starting already with a lower intensity than the other version; a drop of more than 30% occurs between second 1 and 2, followed by a rise until the end of the note. As a result, the version with movement appears to be much smoother than the fermata without movement with its heavy fluctuations.

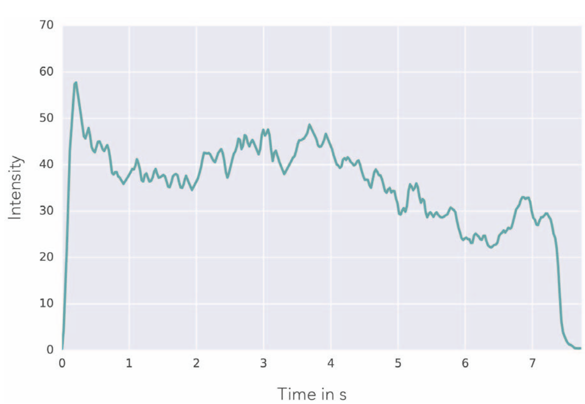


Figure 8. Intensity – Fermata with movement.

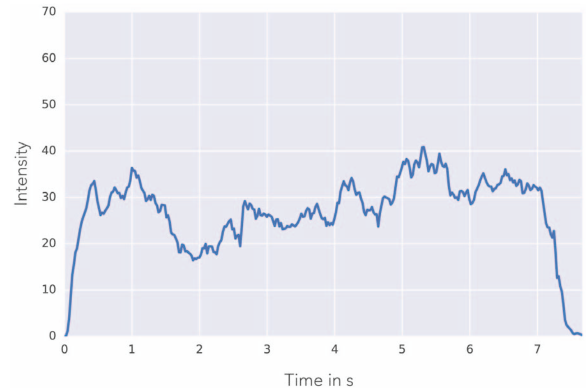


Figure 9. Intensity – Fermata without movement.

V. DISCUSSION

The main purpose of the presented studies was to explore the impact of different gestural archetypes on the musician's body and the sounding result. Firstly, we investigated the reaction of participants on three basic conducting gestures with different shapes of trajectories. The results correspond with Köhler's (1929 & 1947) findings of non-arbitrary mappings between speech sounds and the visual shape of objects and confirm the claim that such correlation also exists between conducting gestures and the elicited tone (Schuldt-Jensen, 2016). Our results for the first time scientifically document the spontaneous and consistent relationship between certain shapes of conducting gestures and the resulting articulation, as well as a significant correlation between size of gesture and evoked volume. Interestingly, different shapes of gestures do not lead to significant difference in terms of loudness of the resulting tone, but the envelope of the evoked tones varies significantly, in that there is a big difference in the intuitively evoked length of sound. The loudness of the note seems to be primarily influenced by the size of the gesture. These findings could be of direct practical relevance, especially for the field of conducting tuition; for example, in view of the fact that there are significant differences in the intuitive reaction to concave versus convex conducting patterns, it would be especially interesting to further analyze teaching books on conducting presenting conducting patterns with a combination of both types.

Another important quality of successful conducting technique is the predictability of the shown gesture. Especially when it comes to delicate musical elements, such as a pizzicato, a subtle unison cue for instruments with differing functionality, and to tempo changes (crucial to musical phrasing), an unpredictably shaped gesture immediately deteriorates the interpretation, because the conductor thus loses control over the timing and the onset-offset-envelope. In order to measure differences in the predictability of conducting gestures, we created tasks with a changing tempo and calculated the onset-delays of the participants' touch-events. While refuting Serrano's (1993) findings, the results confirm Schuldt-Jensen' (2016) analyses, to the effect that convex conducting gestures show the highest predictability whereas mixed and concave gestures show a higher standard deviation especially in sections of deceleration.

The second study aimed at exploring the impact of different conducting gestures on muscle tension of the musicians as well as the resulting sound. The results show significant findings in all investigated categories for both the conductor's muscle tension and the intensity of the evoked sound. However, the outcome has a number of limitations: Firstly, due to the small number of participants, no significance could be reached when measuring the violinists' muscle tension. Secondly, in addition to the individual and slightly varying playing techniques of the participants, every one of them used their own instrument. Both of these factors presumably led to inconsistencies in the measured muscle tension as well as the sound quality. In future studies, the same instrument should be used for all participants, and the individual differences in instrumental technique – if unavoidable – should be documented and be taken into consideration during data analysis.

The reactions on concave, convex and mixed conducting gestures show significant differences in intensity of the evoked tone; interestingly, the amount of tension involved in the production of the conductor's movement seems to be in inverse proportion to the intensity of the resulting tone. Furthermore, trajectories with a natural gravitational pendulum movement, executed through lifting the arm and then releasing the potential energy into the trajectory, such as convex and (partly) mixed, apparently evoke a higher level of intensity than the concave gesture with its bouncing movement against gravity.

The results of the second category (pronation, supination, neutral) show that not only the form of the conducting gesture leads to differences in intensity. Also, different forearm rotations used with the same (convex) form of trajectory influence the intensity of the induced tone significantly. Both of the standard rotations pronation and supination (prescribed as standard positions in many books on conducting) evoke a lower intensity, compared to a neutral hand posture.

As for the fermata, though the two types seem to be rather alike, they nevertheless led to surprisingly distinct differences. Although the movement of the hand during the duration of the fermata needs only slightly more muscle tension on the conductor's side, this action resulted in a much higher overall intensity of the evoked tone and a lower fluctuation of intensity, which combined improves the sound quality consistency of fermatas substantially.

An important implication of our findings is that both the form of the trajectory and the hand posture elicit significant differences in sound quality, especially when the multiplication of the individual results (due to the high number of violinists and other string players in an orchestra) is taken into account. Concerning health related implications, as i.e. overstraining, future studies should measure the impact of different conducting gestures over a longer period of time. As for the conductor's risk of overstraining, by using convex conducting gestures with a neutral hand posture (instead of the other gestural types mentioned above) he can seemingly reduce the required muscle tension on a permanent basis and at the same time evoke a higher intensity of the sound of his ensemble.

VI. CONCLUSION

We presented two studies exploring the physiological impact and the sounding result of different conducting gestures. Our findings show that there are consistent reactions to different types of trajectories and hand postures, indicating that it does indeed matter musically how conducting gestures are formed and executed. However, our findings do not aim to define any of the investigated types of gestures as being right or wrong. But it turns out that since every gesture has a unique sounding consequence, certain gesture types are more capable – more economical and effective – of reaching predefined musical/ interpretational goals than others.

REFERENCES

Bowen, J. A. (2005). The Cambridge Companion to conducting (Reprint). Cambridge: Cambridge University Press.

Clynes, M. (1978). Sentics: The Touch of the Emotions. Garden City, N.Y.: Anchor Press/Doubleday.

Colson, J. F. (2012). Conducting and Rehearsing the Instrumental Music Ensemble: Scenarios, Priorities, Strategies, Essentials, and Repertoire. Scarecrow Press.

Farberman, H. (1997). The art of conducting technique: A new perspective. Miami: Belwin-Mills.

Galkin, E. W. (1988). A history of orchestral conducting in theory and practice. New York: Pendragon Press.

Göstl, R. (2006). Chorleitfaden : motivierende Antworten auf Fragen der Chorleitung. Regensburg: ConBrio.

Green, E. A. H. (1992). The modern conductor : a college text on conducting based on the technical principles of Nicolai Malko as set forth in his The conductor and his baton. Englewood Cliffs, N.J. : Prentice-Hall.

Hattinger, W., (Author). (2013). Der Dirigent: Mythos – Macht – Merkwürdigkeiten. Kassel: Bärenreiter.

Haverkamp, M. (2001). Application of Synesthetic Design as multi-sensory Approach on Sound Quality. Retrieved from http://www.michaelhaverkamp.de/Syn_Wahrn_Ger_update_05_HAV.pdf

Hinton, E. L. (2008). Conducting the wind orchestra.

Jankovic, D. (2006). Development of the takete-maluma phenomenon: Children's performance on choice and visualisation task. PERCEPTION, 35, 90–91.

Köhler, W. (1929). Gestalt psychology. New York, H. Liveright, 1929.

Köhler, W. (1947). Gestalt psychology (2nd ed.). New York, H. Liveright, 1947.

Lawson, C. (2003). The Cambridge companion to the orchestra. Cambridge, U.K. ; New York : Cambridge University Press, 2003.

Lumley, J., & Springthorpe, N. (1989). The art of conducting : a guide to essential skills. London: Rhinegold.

Maiello, A. J., Bullock, J., & Clark, L. (1996). Conducting: A Hands-on Approach. Alfred Music Publishing.

McElheran, B. (2004). Conducting Technique: For Beginners and Professionals. OUP USA.

Nielsen, A. K. S., & Rendall, D. (2013). Parsing the role of consonants versus vowels in the classic Takete-Maluma phenomenon. Canadian Journal Of Experimental Psychology = Revue Canadienne De Psychologie Expérimentale, 67(2), 153–163. http://doi.org/10.1037/a0030553

Oakley, I., & Brewster, S. (Eds.). (2013). Haptic and Audio Interaction Design (Vol. 7989). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved from http://link.springer.com/10.1007/978-3-642-41068-0

Phillips, K. H. (1997). Basic techniques of conducting. Oxford: Oxford University Press.

Schuldt-Jensen, M. (2008). Grundkurs Chorleitung. Edition Continuo,.

Schuldt-Jensen, M. (2016). Sémiotique de la musique. (P. A. Brandt & J. R. do Carmo Jr., Eds.). Liège (Belgique): Presses Universitaires de Liège.

Seaman, C. (2013). Inside conducting. Rochester, NY : University of Rochester Press, 2013.

Serrano, J. G. (1993). Visual perception of simulated conducting motions (A.Mus.D.). Ann Arbor, United States. Retrieved from http://search.proquest.com.libproxy.mit.edu/pqdtglobal/docview/304062873/abstract/9DA93A1082F146BDPQ/1

Sonic Visualiser. (2016, May 2). Retrieved from http://www.sonicvisualiser.org/

Thalmic Labs. (2016, May 2). Retrieved from https://www.myo.com/

Dyadic improvisation and mirroring of finger movements

Tommi Himberg[1], Maija Niinisalo[1] and Riitta Hari[1]
[1]Department of Neuroscience and Biomedical Engineering, Aalto University, Espoo, Finland

Most synchronisation studies investigate rhythmic movements and note onset synchrony, where metrical structure generates strong expectations of timing of future events. We studied kinematics and coordination in a mirror game using fluent, improvised movements that either give rise to a moderate sense of pulse (circle drawing) or are pulse-free (free movement). 32 participants (18 f, 14 m; mean age 25.2 years) took part in dyads, standing face-to-face, pointing their right index fingers at each other, with fingertips 10–15 cm apart. In turn, one of the p's was appointed the leader, or the dyad was instructed to share leadership. Hand movements were recorded with optical motion capture.

Compared to the leader–follower condition, sharing leadership resulted in more synchronous circles, smoother free movements, and strong mutual adaptation. Whereas the "follower" would often perform free movements with jitter (movements at 2–3 Hz range), trying to follow the leader's moves, in the joint leadership tasks this jitter was 23% lower. Such "co-confident motion" was also observed in the 1D, machine-mediated joint leadership mirror task by Noy et al. (2011), but only in experienced improvisers. In contrast, our p's did not have extensive experience in music, dance, or acting, but still easily reached co-confident state in our un-mediated, effectively 2D movement task.

In leader-follower tasks, the lag between participants was app. 0.3 seconds. The joint leadership trials resulted in mutual adaptation, with both p's "following" each other at similar lags, windowed analysis revealing that the direction of the lag varied at sub-second intervals. Importantly, there were no velocity differences between the leadership conditions.

Our free, fluent movement tasks extend the range of current entrainment studies from the strictly rhythmic and metric domain towards fluent movements, which are important for ensemble coordination, and add to the growing literature on movement mirroring.

Let's dance: Examining the differences in beat perception and production between musicians and dancers using motion capture

Tram Nguyen[1,2] & Jessica A. Grahn[1,2]
[1]The Brain and Mind Institute, The University of Western Ontario, London, Ontario, CANADA
[2]Department of Psychology, The University of Western Ontario, London, Ontario, CANADA

Most of us can perceive and produce a regular 'beat' in response to musical rhythms, but individuals vary in how well they can do so. Beat perception and production may be influenced by music (Repp & Doggett, 2007) or dance training (Miura et al., 2011). It may also be influenced by the particular musical instrument or style of dance with which an individual has experience with. For example, training in percussive instruments (e.g., drums or keyboards) and percussive styles of dance (e.g., breakdancing or tap) may lead to better beat processing abilities than training in nonpercussive instruments (e.g., brass or strings) or nonpercussive styles of dance (e.g., ballet or contemporary) (Cameron & Grahn, 2014; Cicchini et al., 2012; Miura et al., 2011). In our previous work, we examined how percussive and nonpercussive music or dance experience influenced beat perception and production, using subtests of the Beat Alignment Test (Gold-MSI, 2014). We found that musicians did better than dancers, but percussive training did not have a significant effect. However, the BAT uses finger tapping to measure beat production; it is possible that the musicians' advantage does not generalize to types of movement more familiar to dancers, such as whole-body movement.

As a follow up, we examined how percussive and nonpercussive music or dance experience influenced beat production when participants bounced to the beat of the music. We used motion capture to assess whole body movements. For the beat perception task, we asked participants to judge whether a bouncing stick figure is on or off the beat. Preliminary data show that the musician advantage persists; musicians were significantly more accurate at both beat perception and production than dancers. Thus, the effects of music experience appear to generalize to beat perception and production assessed with whole-body movement.

Effect of musical training on auditory perception in older adults

E. Dubinsky[1], G. Nespoli[1], L. Smith[1], P. Ammirante[1], & F.A. Russo[1]
[1]Department of Psychology, Ryerson University, Toronto, Ontario, Canada

Hearing loss, which most adults will experience to some degree as they age, is associated with increased risk of cognitive impairment and decreased quality of life. Musicianship has been shown to improve aspects of auditory and cognitive processing, but has not been studied as a short-term intervention for improving these abilities in older adults. The current study investigates whether short-term musical training can improve three aspects of auditory processing: perception of speech in noise, pitch discrimination, and the neural response to brief auditory stimuli (frequency following response; FFR). This study also examines several measures of cognition – working memory and inhibitory control of attention – in order to explore the extent to which gains in auditory perception might be related to improvements in cognitive abilities. Thirty-two older adults (aged 50+) participated in a choir for 10 weeks, during which they took part in group singing (2 hours/week) supported by individual online musical training (1 hour/week). Choir participants (n=32) and an age-matched control group (n=20) underwent pre- and post-training assessments, conducted during the first week of the choir and again after the last week. Cognitive and auditory assessments were administered electronically, and the FFR was obtained using electroencephalography (EEG). Preliminary statistical analyses showed that choir participants improved across all auditory measures, while the control group showed no differences. As well, choir participation resulted in higher scores on a non-auditory attention task, suggesting that musical training can lead to cognitive changes outside the auditory domain. A correlational analysis demonstrated that older adults with increased hearing loss achieved greater gains as a result of this intervention. These findings support our hypothesis that short-term choir participation is an effective intervention for perceptual and cognitive aspects of age-related hearing loss.

Experience of sound making changes neuromagnetic responses in perception

Bernhard Ross[1,2], Masih Barat[1], and Takako Fujioka[3]
[1] Rotman Research Institute, Baycrest Centre, Toronto, ON, Canada
[2] University of Toronto, Department for Medical Biophysics, Toronto, ON, Canada
[3] Center for Computer Research in Music and Acoustics, Stanford University, Stanford, CA, U.S.A.

Brain processes underlying sound perception are different when passively listening or being actively involved in sound making such as in music. During the action-perception cycle of sound making, auditory and sensorimotor areas interact closely, thus integrating multimodal sensory feedbacks into subsequent action planning and execution. Auditory evoked responses during sound making such as vocalization are attenuated compared to passively listening to same recorded sound. Similar movement-related response suppression had been found for sound making actions with instruments.

In addition to showing suppressed auditory brain responses while making sound, our hypothesis was that newly engaged action-perception connections would affect how we interpret the sound during subsequent listening.

Eighteen young healthy adults participated in three magnetoencephalography (MEG) sessions while (1) passively listening to the recorded sound of a chime, (2) actively making sounds on a chime, and (3) listening again to the recorded sounds.

In MEG data analysis, we localized sources of auditory brain responses in the bilateral auditory cortices, and projected the MEG signals into time series of left and right auditory cortex activity.

Averaged auditory evoked responses recorded during passive listening showed typical P1, N1, and P2 peaks at 50 ms, 100 ms and 200 ms after sound onset and a sustained wave according to the duration of the lingering chime sound. During sound making, we observed similar P2 and sustained responses, while the N1 response was substantially attenuated. Comparison of brain activity during naive and experienced listening showed an increase in the P2 response and emerging beta activity after sound making.

Previous studies found that the P2 response and beta oscillations are specific to perception rather sensation of sound. Therefore, we discuss the experimental results as objective indicator of short-term experience induced neuroplastic changes of perception.

ISBN 1-876346-65-5 © ICMPC14

Brainstem and cortical responses to sounds presented on the strong and weak beats

Kyung Myun Lee[1], Kyogu Lee[2] and Il Joon Moon[3]
[1]School of Humanities & Social Sciences, Korea Advanced Institute of Science and Technology, Daejeon, South Korea
[2]Graduate School of Convergence Science and Technology, Seoul National University, Seoul, South Korea
[3]Department of Otorhinolaryngology-Head & Neck Surgery, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, South Korea

According to dynamic attending theory, attention allocated to metrically strong beats facilitates auditory perception of sounds. Enhanced cortical auditory processing at times of high metric strength was evidenced by heightened N1 and P2 peaks in ERP research. However, very little is known about how the subcortical processing of sounds is influenced by rhythm perception. A previous study showed that the auditory brainstem responses (ABRs) were enhanced when sounds were aligned with the beat of music, compared to shifted away from the beat, but it remains to be explored how the subcortical processing of sounds is different among various metrical positions. By measuring and comparing both subcortical and cortical responses to four different beats of the quadruple meter, this study examined how early auditory processing could be modulated by temporal attention driven by the quadruple meter. In order to prime the quadruple meter, a sinusoidal four-tone sequence composed of A7 (3520Hz), A6 (1760Hz), A6 (1760Hz), and A6 (1760Hz) (500ms IOI) was repeatedly played, while a short speech sound, /da/, was simultaneously presented every 500ms. ABRs and P1 responses to the speech sound, /da/, were measured and compared among four different beat positions. The result showed that the onset latencies and amplitudes of ABRs were not significantly different among four beats, whereas P1 was significantly reduced on the strong beat. The results indicate that the effects of temporal attention guided by the quadruple meter is robust only on the cortical level and the details of metric strength is not reflected on the subcortical processing of sounds.

Auditory Brain Oscillations during the Perception and Imagery of Musical Pitch

Rebecca W Gelding[1,2], William F Thompson[2,3], Blake W Johnson[1,3]
[1]Department of Cognitive Science, Macquarie University, Sydney, NSW, Australia
[2]ARC Centre of Excellence in Cognition and Its Disorders, Macquarie University, Sydney, NSW, Australia
[3]Department of Psychology, Macquarie University, Sydney, NSW, Australia

The Pitch Imagery Arrow Task (PIAT) was designed to reliably induce pitch imagery in individuals with a range of musical training. The present study used the PIAT to investigate the effects of pitch imagery on intrinsic brain rhythms measured with magnetoencephalography (MEG). Participants were 19 healthy adults whose musical expertise ranged from novice to professional musicians. Participants performed the pitch imagery task with a mean accuracy of 80.2% (SD = 14.9%). Brain sources were modeled with bilateral auditory cortical dipoles. Time-frequency analyses revealed prominent experimental effects on auditory cortical beta-band (15-30 Hz) rhythms, which were more strongly desynchronized by musical imagery than by musical perception. Such effects in the beta band may reflect a mechanism for coordinating auditory and motor processing of temporal structure in music. In addition, cortical oscillations showed a marked relationship to a psychometric measure of musical imagery. Specifically, the magnitude of cortical oscillations in the high gamma band (60-80 Hz) was significantly and inversely related to auditory manipulation scores measured by the Bucknell Auditory Imagery Scale Control (BAIS_C) which evaluates the degree of mental control that individuals perceive they have over their imagery. Gamma band activity reflects the recruitment of neuronal resources in auditory cortex, and the observed negative correlation is consistent with the interpretation that pitch imagery is a less effortful process for those who self-report higher mental control. These results add to a growing body of evidence that intrinsic brain oscillations play crucial roles in supporting our ability to mentally replay or create music through musical imagery.

Metastability and effective brain connectivity of groove music

Maria A.G. Witek[1], Eric F. Clarke[2], Mikkel Wallentin[3,4], Mads Hansen[6], Matthieu Gilson[5], Gustavo Deco[5], Morten L. Kringelbach[1,3,7] & Peter Vuust[1,3]
[1]Center for Music in the Brain, Dept. of Clinical Medicine, Aarhus University & Royal Academy of Music, Aarhus/Aalborg, Denmark
[2]Faculty of Music, University of Oxford, UK
[3]Center of Functionally Integrative Neuroscience, Aarhus University, Denmark
[4]Department of Linguistics, Cognitive Science and Semiotics, Aarhus University, Denmark
[5]Department of Information and Communication Technologies, Universitat Pompeu Fabra, Spain
[6]Department of Psychology and Behavioural Sciences, Aarhus University, Denmark
[7]Department of Psychiatry, University of Oxford, UK

Groove is a musical quality associated with a pleasurable desire to move. Phenomenologically, the ‘groove state of listening’ is described as involving feelings of entrainment, stability and temporal equilibrium; in many ways similar to the optimal metastability found in some dynamical systems. We investigated this link between a subjective state and brain networks using neuroimaging and whole-brain computational modelling of the brain’s temporal dynamics during listening to groove music. We used stimuli from our previous rating-survey, where we found that syncopation – a rhythmic structure that stimulates metric expectations – is an important component in groove. Specifically, intermediate levels of syncopation elicit the most wanting to move and the most pleasure with groove music. We scanned the brains of twenty-six healthy participants using functional Magnetic Resonance Imaging. Whole-brain computational modelling was used to measure the changes in metastability and effective connectivity when listening to drum-breaks with low, medium and high degrees of syncopation. We were able to characterize the temporal dynamics and directionality of brain activity during listening to groove music. We hypothesised that groove music would be associated with networks involved in reward, auditory and motor functioning. Specifically, we were interested in whether medium degrees of syncopation – which are the most pleasure- and movement-inducing in groove – would be associated with optimality in metastability. In other words, to what extent is the temporal stability that listeners experience with groove reflected in the coordinated oscillations of the interacting brain networks? The novel whole-brain computational modelling employed in this study is a significant improvement of previous correlational neuroimaging studies, and as such our study provides a significant breakthrough not only in the neuroscience of music but also in brain connectivity more broadly.

Neural and physiological correlates of rhythm induced trance

Michael J. Hove[1,2], Assal Habibi[3], Johannes Stelzer[4], and B. Rael Cahn[3,5]
[1]Psychological Science, Fitchburg State University, USA
[2]Department of Psychiatry, Harvard Medical School, USA
[3]Brain and Creativity Institute, University of Southern California, USA
[4]Max Planck Institute for Biological Cybernetics, Germany
[5]Department of Psychiatry, Keck School of Medicine, University of Southern California, USA

Rhythmic drumming has been used for centuries to alter consciousness and induce states of trance. Rhythm-induced trance is commonly practiced in shamanism, humanity's most ancient spiritual and healing tradition. Most forms of shamanism use similar repetitive rhythms, which suggests a common biological basis. Despite similar techniques across cultures and powerful phenomenological content, little is known about the neural and physiological mechanisms underlying this form of trance. We present a series of studies that examine the neural and physiological correlates of rhythm-induced trance in experienced shamanic practitioners. In the first study, we used fMRI to examine the neural patterns associated with trance. Experienced shamanic practitioners (n=15) underwent 8 minute brain scans while they listened to rhythmic drumming and entered a trance state (or remained in non-trance in a control condition). During trance, brain networks displayed notable reconfigurations, including increased connectivity in regions associated with internal thought (the default mode's posterior cingulate cortex) and cognitive control (dorsal anterior cingulate cortex and insula), as well as decreased connectivity within the auditory pathway. Together this network configuration suggests perceptual decoupling and that the repetitive drumming was gated out to maintain an internally oriented stream of consciousness. We followed up this work in an EEG/ERP study that used a similar design to examine auditory gating and network activity while shamanic practitioners (n=16) experienced rhythm-induced trance and a control mind-wandering state. We will also present work that examines physiological correlates and responses to trance including a study on heart-rate variability (a measure of sympathetic vs. parasympathetic dominant states) (n=14) and the stress response measured in terms of cortisol levels after a ritual session (n=20). Together this work helps explicate the common use of rhythmic drumming to alter consciousness and why trance is a valuable tool to facilitate insight in many cultures.

A Functional and Effective Connectivity Examination of the Brain's Rhythm Network

Dane E. Anderson[1], Chris L. Keown[1,2], Kathleen Gates[3], Aniruddh Patel[4], Daniel Callan[5,6], Ralph-Axel Müller[1,7], John Iversen[8]
[1]Brain Development Imaging Laboratory, San Diego State University, USA
[2]Department of Cognitive Science, University of California San Diego, USA
[3]Department of Psychology, University of North Carolina at Chapel Hill, USA
[4]Department of Psychology, Tufts University, USA
[5]Center for Information and Neural Networks (CiNet), Osaka, Japan
[6]National Institute of Information and Communications Technology (NICT), Osaka University, Japan
[7]Department of Psychology, San Diego State University, USA
[8]Swartz Center for Computational Neuroscience, University of California San Diego, USA

Deficits in the ability to perceive and produce a beat correlate with the impairment of coordinated movement in disorders such Parkinson's disease. Neuroimaging has shown that rhythmic movements and maintaining a steady beat involve the interaction of auditory and motor areas, within a wider network of prefrontal, parietal, cerebellar, and striatal regions. How these dynamic interactions establish an internal sense of beat and enable us to produce rhythms is only beginning to be understood, yet may ultimately help develop more effective interventions for movement disorders. This study therefore examined brain activity across a large network using functional (fMRI), functional connectivity (fcMRI), and effective connectivity MRI (ecMRI) techniques during a beat production and maintenance task of contrasting strongly and weakly beat-inducing stimuli. fMRI BOLD activation results from14 participants (4F, 24 ± 4.5yrs) are consistent with the previous literature. Tapping to a strong beat compared to self-paced tapping showed the greatest relative activation in the bilateral auditory cortices, supplementary motor areas (SMA), premotor cortices (PMC), and inferior frontal gyri (IFG). Relative to a weak beat, significant activation was seen in the putamen and thalamus. Functional connectivity was examined across a more comprehensive group of brain regions of interest (ROIs) then has been previously reported (18 ROIs), including regions the cerebellum, thalamus, inferior parietal lobe, and caudate nucleus. These analyses demonstrated that connectivity is remarkably stable regardless of stimulus beat strength or self-paced tapping. Examination of the strongest signal correlations showed the left IFG, thalami, and particularly the SMA may act as important connectivity "hubs." A preliminary ecMRI analysis using Group Iterative Multi Model Estimation (GIMME) confirmed the fcMRI analyses' outline of the network's structure as well as its identification of important network hubs.

Exploring music & language processing and emotions with ECoG

Organizers: C.A. Mikutta[1,2,3], D. Omigie[4] & R.T. Knight[1]
Discussant: P. Janata[5]
[1]Helen Wills Institute of Neuroscience, and the Department of Psychology, University of California Berkeley, 132 Barker Hall, Berkeley, CA 9472 USA
[2]Department of Neurology, Inselspital and University of Berne, Switzerland
[3]Department of Psychiatric Neurophysiology, University Hospital of Psychiatry and University of Bern, Bern, Switzerland
[4]Music department, Max Planck Institute for Empirical Aesthetics, Germany
[5]Department of Psychology, and Center for Mind and Brain, University of California, Davis, CA, USA

Electrocorticography (ECoG), which employs direct cortical recordings from either subdural or depth electrodes, provides a unique method to study cognition and emotion in humans.

Clinically ECoG is the gold standard for locating epileptic sources in therapy refractory epilepsy patients. Within cognitive neuroscience it provides the combination of high temporal (in ms range) and high spatial (in mm range) resolution. Those features offer a high potential for (a) exploring the neural encoding of music within the secondary auditory areas via cortical surface electrodes and (b) exploring activations of limbic structures due to music induced emotions via depth electrodes in the hippocampus and amygdala.

Prof. Knight will focus on the use of auditory stimuli to study language processes using ECoG. He will address the cortical mechanisms of phoneme and word representation, categorical phoneme representation and speech production.

Dr. Mikutta will present ECoG data recorded from 28 patients while they listened to music. Using the high gamma filtered data from the informative electrodes (144 out of 1760 over all patients) a spectrotemporal decoding model was fitted. This model was able to reconstruct the spectrogram properties of a given music directly from cortical recordings of the brain. The perceptual quality of the sound reconstructions increased with the spatial density of the electrodes, and the extent of frequency tuning sampling. These results reveal that the spectro-temporal sound representation of music can be decoded using neural signals acquired directly from the human cortex.

Finally Dr. Omigie will demonstrate the different roles of hippocampus and amygdala within the feeling of expectedness and unexpectedness in music. She presents data from patients with depth electrodes in the hippocampus, amygdala and lateral temporal areas. Using Granger causality analysis she demonstrates differences in information flow during expected and unexpected notes in a melody. While the hippocampus processes subjective feelings of familiarity in response to expected notes, the amygdala may facilitate feelings of tension or surprise in response to unexpected high IC ones.

Electrocorticography and Language

Robert T. Knight[1]
[1]Helen Wills Institute of Neuroscience, and the Department of Psychology, University of California Berkeley, 132 Barker Hall, Berkeley, CA 9472 USA

Language, attention, perception, memory, motor control and executive control generate focal high frequency oscillatory activity in the range of 70-250 Hz (high gamma, HG). The HG response in the human electrocorticogram (ECoG) precisely tracks auditory processing in the neocortex and can be used to study the spatio-temporal dynamics of auditory and linguistic processing. ECoG data will be reviewed addressing the neural mechanisms of speech suppression, categorical representation and the timing of speech perception and production in peri-sylvian language regions.

Reconstruction of music from direct cortical recordings in humans

C.A. Mikutta[1], S. Martin[1,2], G. Schalk[3,4], P. Brunner[3,4], R.T. Knight[1], & B.N. Pasley[1]
[1]Helen Wills Institute of Neuroscience, and the Department of Psychology, University of California Berkeley, 132 Barker Hall, Berkeley, CA 9472 USA
[2]Department of Bioengineering, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
[3]New York State Department of Health, Wadsworth Center, Albany, NY, USA
[4]Department of Neurology, Albany Medical College, Albany, NY, USA

Music is a universal aspect of human behavior spanning all cultures. We addressed how music is represented in the human brain by recording electrocorticographic signals (ECoG) from 28 patients with intracranial electrodes covering left and right auditory cortices. A decoding model based on sound spectral features and high gamma cortical surface potentials was able to reconstruct music with high sound quality as judged by human listeners. The perceptual quality of the sound reconstructions increased with the spatial density of the electrodes, and the extent of frequency tuning sampling. These results reveal that the spectro-temporal sound representation of music can be decoded using neural signals acquired directly from the human cortex.

Hippocampus and Amygdala interactions associated with Musical Expectancy

Diana Omigie[1], Marcus Pearce[2], Severine Samson[3]
[1]Music department, Max Planck Institute for Empirical Aesthetics, Germany
[2]Music Cognition Lab, Queen Mary, University of London, UK
[3]Neuropsychology and auditory cognition lab, University of Lille, France

Numerous studies have linked music-induced emotions to activity in the amygdala (AMYG) and hippocampus (HIPP). However, how these regions respond to (expectations about) unfolding musical structure has seen relatively little research. This is unfortunate, given that musical pitch expectancy is held to be a rich source of musical emotions. To fill this gap, the current study explored the potential role of the HIPP and AMYG in mediating musical expectancies and thereby emotions. Electrical activity was recorded from epileptic patients implanted with depth electrodes while they listened to melodies comprising notes characterized in terms of Information content (IC, low IC= Expected, high IC=Unexpected). Responses to low and high IC notes were compared in terms of Granger Causality (GC) and cross-regional Phase Amplitude Coupling (PAC). GC revealed a more hub-like behavior of HIPP regions during low IC notes. Further, in line with hierarchical models that emphasize frequency specificity of bottom-up (BU) and top-down (TD) interactions: 1) TD flow (HIPP to lateral temporal (LT)) and BU flow (LT to HIPP) tended to be in the theta and gamma range, respectively and 2) PAC revealed HIPP theta phase modulation of beta/gamma power in the LT areas but not the reverse. Finally, in contrast to the HIPP, and in line with the notion of AMYG as a salience detector, amygdala showed increased connectivity (notably, with the HIPP) during the processing of high IC relative to low IC notes. In conclusion, HIPP and AMYG may have different but complementary roles in the processing of (expectations about) musical pitch structure. While HIPP may facilitate subjective feelings of "HOME" or familiarity in response to expected/ low IC notes, the amygdala may facilitate feelings of tension or surprise in response to unexpected high IC ones. The continually changing dynamics of these two regions likely comprise an important neural substrate of everyday music-induced emotions.

Neural Correlates of Auditory and Language Development in Children Engaged in Music Training

Assal Habibi[1], B. Rael Cahn[1,2], Antonio Damasio[1] and Hanna Damasio[1]
[1]Brain and Creativity Institute, Dornsife College of Letters, Arts and Sciences, University of Southern California, USA
[2]Department of Psychiatry , Keck School of Medicine, University of Southern California, USA

Several studies comparing adult musicians and non-musicians have shown that music training is associated with functional and anatomical brain differences. It is unknown, however, whether those differences result from lengthy musical training or from pre-existing biological traits or social factors favoring musicality. As part of an ongoing 5-year longitudinal study, we investigated, using behavioral, neuroimaging and electrophysiological measures, the effects of a music training program on the auditory and language development of children, over the course of three years beginning at 6-7. The training was group-based and inspired by El Sistema. We compared the children in the music group with two groups of "control" children of the same socio-economic background, one involved in sports training, another not involved in any systematic training. Prior to participating, children who began training in music were not different from those in the control groups relative to cognitive, motor, musical, or brain measures. After three years, we now observe that children in the music group, but not in the two control groups, show an enhanced ability in pitch and rhythm perception and phonological processing and an accelerated maturity of auditory processing as measured by cortical auditory evoked potentials. Our results support the notion that music training results in significant brain changes in school aged children.

Music use predicts neurobiological indices of well-being and emotion regulation capacity

Tan-Chyuan Chin[1] and Nikki Rickard[1,2]
[1]Melbourne Graduate School of Education, The University of Melbourne, Victoria, Australia
[2]School of Psychological Sciences, Monash University, Victoria, Australia

Music engagement for the purposes of cognitive and emotional regulation has been shown to be consistently associated with, and predictive of, subjective indices of well-being and adaptive emotion regulation strategy. The present study investigated the extent to which music engagement contributed to neurophysiological substrates of well-being, after statistically controlling first, for age and gender, and secondly, music listening and training in a sample of 38 participants between the ages of 18 and 42 (M = 27.11, SD = 6.66). Frontal asymmetry and heart rate variability were acquired using electroencephalography (EEG) and electrocardiography (ECG). Intentional use of music for cognitive and emotional regulation predicted both well-being and emotion regulation capacity, as measured respectively by resting frontal asymmetry and high frequency heart rate variability. Furthermore, this predictive value was maintained after statistically controlling for age, gender, music listening and music training. These findings suggest that purposeful engagement with music for cognitive and emotional regulation could be strategically targeted for future well-being interventions.

Neural Correlates of Dispositional Empathy in the Perception of Isolated Timbres

Zachary Wallmark,[*1] Choi Deblieck,[#2] Marco Iacoboni[*3]

[*] *SMU Meadows School of the Arts, Dallas, TX, USA*

[#] *Academic Center for ECT and Neuromodulation, University Psychiatric Center, University of Leuven, BELGIUM*

* *Ahmanson-Lovelace Brain Mapping Center, University of California, Los Angeles, USA*

[1]zwallmark@smu.edu, [2]choi.deblieck@gmail.com, [3]iacoboni@ucla.edu

ABSTRACT

Neuroimaging studies on sensorimotor integration and social cognition suggest that individuals with high dispositional empathy (measured by the Interpersonal Reactivity Index, or IRI) process human-made action sounds with greater involvement from motor areas than individuals with lower dispositional empathy. Does high empathy correlate with increased motor activity when listening to musical sounds as well? To address this question, we performed an fMRI study with fifteen subjects. Stimuli consisted of 12 isolated 2-second natural instrument and vocal timbres in "regular" and "noisy" versions. With "empathic concern" scores from the IRI included as covariates, we found that high-empathy individuals processed noisier qualities of timbre with heightened motor activity relative to regular timbres, particularly in the supplementary motor area (SMA). We interpret these results through the lens of embodied music cognition to suggest that dispositional empathy modulates motor resonance during the perception of the sounds of others, especially when these sounds signal heightened arousal and negative valence.

I. INTRODUCTION

Music is a portal into the interior lives of others. As a "virtual social agent" that sonically discloses the affective and cognitive states of actual or imagined human actors, musical engagement can function as a mediated form of social encounter, even when listening alone (Leman, 2007). The concept of empathy has generated sustained interest in recent years among music cognition scholars seeking to better account for the social valence of musical experience and behaviors; however, the precise relationship between music processing and empathy remains poorly understood. Empathy modulates our engagement with other people: does it modulate music processing as well? Do empathic individuals process music differently than lower-empathy people?

If we consider music through a social-cognitive lens, it would seem plausible that those with a greater capacity to empathize with other people might also respond to music-as-social-stimulus differently on a neurophysiological level. Individuals with a high level of *dispositional empathy*—i.e., empathy as a self-reported character trait—have been shown to process non-musical action sounds with a greater degree of involvement from motor areas (Gazzola, Aziz-Zadeh, & Keysers, 2006; Kaplan & Iacoboni, 2006). Furthermore, such simple mechanisms of "motor resonance" or "emotional contagion" (Hatfield, Cacioppo, & Rapson, 1994) appear to underpin more complex, deliberative acts of prosocial decision-making (Christov-Moore & Iacoboni, 2016).

To explore the neural substrates underlying the relationship between empathy and music perception, we carried out a study using functional magnetic resonance imaging (fMRI). In the present study we focused on a single attribute of musical sound—timbre—to investigate the effects of empathy on how listeners process isolated vocal and instrumental sounds. We developed two main hypotheses: First, we anticipated that empathy would be correlated with motor resonance; second, following an embodied cognitive model, we hypothesized that subjectively and acoustically "noisy" timbral qualities would preferentially engage similar motor regions among subjects with high IRI scores.

II. METHODS

A. Subjects

Fifteen participants (8 female, 18–20 years old) with a range of musical backgrounds were recruited, all right-handed, normal or corrected-to-normal vision, and no history of neuropsychiatric disorder. The study was approved by the UCLA Institutional Review Board.

B. Stimuli

We recorded 12 2s signals: 3 electric guitar, 3 tenor saxophone, 3 *shakuhachi* flute, and 3 female vocals. For each sound producer, signals were divided into *regular* and *noisy* versions, defined with the musicians' assistance according to the specificities of each source: (1) regular condition, (2) noisy condition #1, and (3) noisy condition #2. Signals were the same pitch (233 Hz, B-flat 3) and were manually equalized for loudness (for details see Wallmark, 2014).

C. Perceptual Attributes of Stimuli

To verify perceptual differences between regular and noisy versions, we conducted a pilot experiment (N = 36) asking subjects to rate the signals on five perceptual features using a numbered horizontal rating scale with bipolar labels consistent with a VAME-modified semantic differential paradigm (Kendall & Carterette, 1993; Osgood, Suci, and Tannenbaum, 1957): (1) bodily exertion required to produce each sound, (2) valence, (3) perceived "brightness," (4) perceived "noisiness," and (5) emotion conveyed from a list of 5 primary emotions (*happiness, sadness, anger, fear,* and *tenderness*). ("Brightness" was omitted from the analysis.) A three-way repeated measures ANOVA with 2 levels for the Perception of Noisiness factor (negative valence and "noisiness"), 4 for the Instrument factor, and 2 ordinal levels for the Noisy Timbre factor (*regular* and *noisy* signals) revealed significant main effects for Instrument $F(3, 27) = 3.88$, $p = .02$, and Noisy Timbre $F(1, 9) = 10.97$, $p < .01$. The difference between the two perceptual measures of noisiness was not significant, $F(1, 9) = 0.34$, $p = 0.58$.

Noisy versions were correlated at a significant level ($p < .05$) with perceived bodily exertion ($r = .55$), negative valence (.55), and *anger* (.87).

D. Acoustical Attributes of Stimuli

Noisy signals were characterized acoustically by elevated inharmonicity, spectral centroid, spectral flatness, zero-cross rate, and auditory roughness, as measured using MIRtoolbox 1.4 (Lartillot & Toiviainen, 2007) in MATLAB.

E. Interpersonal Reactivity Index

The Interpersonal Reactivity Index (IRI), a 28-item survey designed to measure individual differences in empathy, was given to all subjects prior to the scan (Davis, 1983). The index uses 5-point Likert questions to capture four subscales of empathy: *perspective taking, fantasy, empathic concern,* and *personal distress*, of which "empathic concern" (EC) is the most relevant to the present study. The EC subscale attempts to capture the emotional component of empathy, or "feeling with someone else" (Preston & de Waal, 2002). The highest EC score possible is 35 (7 questions x 5 points each); mean score of the 14 participants (one subject did not complete the IRI) was 29.07, *SD* = 3.08.

F. MRI Procedure

Subjects listened to the randomized stimuli while being scanned. Each stimulus was repeated 6 times with 10 ms of silence between each onset (approximately 12 s total per timbre). After the presentation of all 12 stimuli, subjects were given a 16 s baseline period of silence. The full block was repeated 3 times.

G. Data Acquisition, Preprocessing, and Statistics

Images were acquired on a Siemens 3T Trio MRI scanner, and audio stimuli were timed and presented with Presentation software through noise-cancelling, magnet-compatible SereneSound headphones. Image preprocessing and data analysis were performed with FSL version 5.0.4. Images were realigned to the middle volume to compensate for any head motion using MCFLIRT (Jenkinson et al., 2002). Statistical analyses were performed at the single-subject level using a general linear model with fMRI Expert Analysis Tool (FEAT, version 6.00), and group-level analysis was carried out using FSL FLAME stage 1 and 2 (Beckman, Jenkinson, & Smith, 2003). All images were thresholded at $Z > 2.3$ ($p < .01$), corrected for multiple comparisons using cluster-based Gaussian random field theory controlling family-wise error across the whole brain at $p < .05$. A region of interest (ROI) was determined from the task > baseline contrast, and subsequent contrasts were masked for this ROI. In our contrasts, the "noisy" version of the stimuli comprised an equal contribution of noisy conditions #1 and 2.

III. RESULTS

With "empathic concern" (EC) scores added as a covariate, we found activity in a wide range of brain regions during the timbre listening task, including bilateral supplementary motor area (SMA), premotor cortex, and insula (see Table 1).

Table 1. Neural Correlates of Empathic Concern in Timbre Perception

Contrast and regions	MNI coordinates x	y	z	Max. *Z*
Task (timbre) > baseline (silence)				
L SMA	–6	–2	74	3.35
R SMA	8	–2	72	3.54
L superior parietal lobule	–14	–56	64	3.07
R superior parietal lobule	12	–58	62	3.88
L premotor cortex	–30	–8	68	3.57
R premotor cortex	26	–14	62	2.97
L inferior parietal lobule	–54	–26	30	3.11
R inferior parietal lobule	54	–16	36	3.33
L primary somatosensory cor.	–40	–34	50	3.36
R primary somatosensory cor.	38	–34	50	2.82
R Broca's area	52	16	32	2.81
L insula	–36	20	–2	3.22
R insula	42	10	–2	3.27
L putamen	–20	10	–12	3.29
L visual cortex (V2)	–4	–92	–6	3.06
Cerebellum (vermis VIIIa)	0	–66	–34	3.11
Noisy > regular isolated timbres				
L SMA	–10	–26	72	3.32
R SMA	8	–26	68	3.15
Noisy voice > regular voice				
L SMA	–6	–20	66	3.25
R SMA	16	–26	76	2.80
L parietal operculum 1	–46	–30	16	3.21
L Rolandic operculum	–54	–6	12	2.56
L inferior parietal lobule	–58	–20	36	3.14
L Broca's area	–52	28	6	3.34
Noisy saxophone > regular sax				
Paracingulate gyrus	–2	42	22	3.46
Noisy electric guitar > regular gtr.				
L SMA	–16	–24	72	3.02
L superior parietal lobule	–14	–52	64	3.00

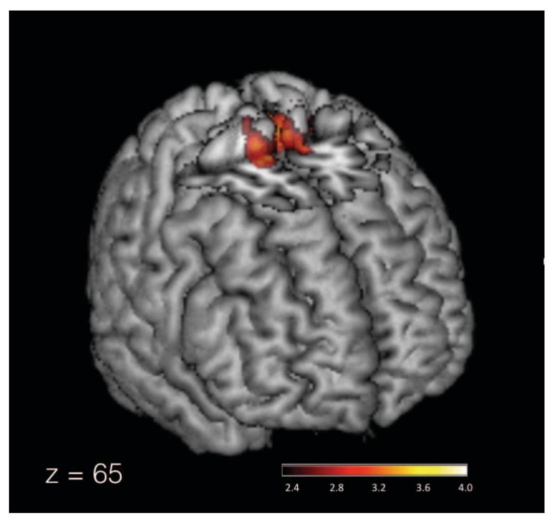


Figure 1. Empathy modulates SMA: "Empathic concern" (EC) covariant in the contrast noisy > regular timbre

In aggregate (i.e., with instrument factor collapsed), perceptually and acoustically "noisy" timbres modulated activity in bilateral SMA to a significantly greater degree among individuals with high dispositional empathy, as shown in Figure 1 ($p < .01$). To make sure that the correlation was not driven by outliers, we extracted parameter estimates from the peak SMA voxel, and plotted the correlations with EC ($r = .76$) as shown in Figure 2.

The electric guitar and the voice exhibited a similar motor pattern. Activity in SMA and superior parietal lobule correlated with empathic concern when subjects heard distorted guitar timbre compared to a regular, "clean" electric guitar signal. Noisy vocal timbres also modulated SMA; additionally, the contrast noisy > regular voice showed correlations with EC in left parietal operculum, Rolandic operculum, inferior parietal lobule, and Broca's area.

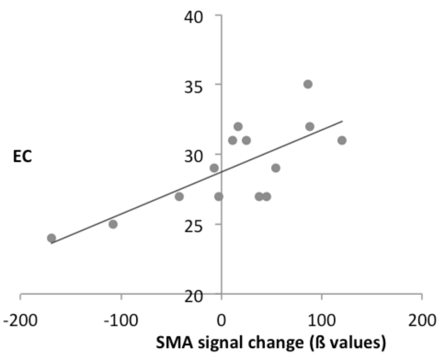


Figure 2. Parameter estimates of peak SMA voxel (*x*) correlated with "empathic concern" (EC) scores (*y*) when subjects hear noisy timbres

IV. DISCUSSION

This study examined the neural correlates of dispositional empathy in the processing of isolated musical timbres, specifically those heard as "noisy." We present two novel findings. First, our study reveals a possible motor component in timbre processing, especially among listeners with higher levels of dispositional empathy. Previous neurophysiological research has demonstrated the involvement of large areas of the temporal lobe (e.g., superior temporal gyrus, superior temporal sulcus, and Heschl's gyrus) in the perception of timbre (Platel et al., 1997; Pantev et al., 2001; Menon et al., 2002; Alluri et al., 2012); but only one study, to our knowledge, has implicated motor regions (Halpern et al., 2004). Our basic task > baseline contrast (without empathy covariate), which was used as an ROI for all subsequent analyses, found the contribution of a range of motor areas in timbre perception (Wallmark, Deblieck, & Iacoboni, in prep), suggesting multi-modal processing of musical timbre. Cross-fertilizing this result with behavioral data from the IRI indicates that this motor resonance is even more acute for empathic people.

Second, the present study suggests that the supplementary motor area (SMA) is preferentially engaged among high-empathy individuals in the perception of subjectively and acoustically "noisy" musical timbres (over "regular" timbral qualities performed on the same sound producers). SMA has been implicated in internally generated movement, coordination of action sequences, and postural stability (Roland et al., 1980; Nguyen, Breakspear, & Cunnington, 2014); it has also been found to contribute to the vividness of auditory imagery, including timbral imagery (Halpern et al., 2004; Lima et al., 2015). Halpern et al. (2004) attributed SMA activity in part to subvocalization of timbral attributes, and the present study would seem to partially corroborate this explanation. Indeed, we interpret this result as a possible instance of sensorimotor integration: SMA activity could reflect a neurophysiological propensity to link sounds with their associated actions.

In a single-neuron study of action execution and observation, the SMA was shown to possess "mirroring" properties; that is, it was active both when subjects performed *and* observed certain hand grasping actions and emotional facial expressions (Mukamel et al., 2010). Even in passive listening, then, timbres may suggest the motions (and emotions) that contribute to them: perhaps the prevalence of SMA here is evidence for the phenomenon of motor resonance demonstrated previously in other auditory domains (Aglioti & Pazzaglia, 2010), including music perception (Zatorre, Chen, & Penhune, 2007). Following this embodied interpretation, people do not just passively listen to different timbres—they enact some of the underlying physical determinants of sound production, whether through subvocalization, biography-specific act/sound associations (Margulis et al., 2009), or other mechanisms of audio-motor coupling. The voice appears to be categorically distinct in this regard: in addition to SMA, we found activation in Broca's area and Rolandic operculum, important areas for speech production and larynx control (Brown, Ngan, & Liotti, 2008), when subjects heard noisy vocal timbres.

What accounts for the difference in neural processes between noisy and regular qualities of sound? Noisy timbre—i.e., spectral characteristics that are generally disliked, such as excessive inharmonicity, spectral centroid, and auditory roughness—appears to modulate motor resonance to a greater degree than non-noisy timbres, and we make sense of this finding through an ecological and embodied lens. Distorted qualities of timbre in vocal and (non-synthetic) instrumental sound production typically index physical exertion, and are associated with high-potency affect and low valence (Scherer, 1995). Via the ecological contingencies of sound production and perception, the listener parses these acoustic cues into their relevant sensorimotor and affective components (Juslin & Västfjäll, 2008; Wallmark, 2014). The ethological association between noisy, aperiodic qualities of timbre and arousal appears to be phylogenetically ancient, and is not exclusive to humans (Tsai et al., 2010; Blumstein, Bryant, & Kaye, 2012). In sum, listeners may respond with greater intensity to more "intense" timbres owing to the ecological urgency typically signaled by such sound qualities. As our results indicate, however, motor resonance is not necessarily dictated by timbral quality alone: there is a probable social-cognitive variable as well (Overy & Molnar-Szakacs, 2009; Launay, 2015).

According to Theodor Lipps's early formulation (*Einfühlung*), empathy enables one individual to intersubjectively "feel into" another through inner imitation. To experience another's point of view, in this account, is not an abstract mental representation of what it feels like "in the other's shoes," but rather constitutes the dynamic co-experience of affective and cognitive states (Gallese, 2007). In musical terms, empathy has been shown to moderate between recognized and induced emotions in music: the greater the dispositional empathic concern on a social level, the more likely an individual is to exhibit a strong affective response to music (Egermann & McAdams, 2013). Though we did not explicitly address musical affect in this study, our results indicate that empathic concern modulates motor resonance, which has been theorized to play an important role in emotional contagion (Molnar-Szakacs & Overy, 2006). Individuals with higher dispositional empathy may be more likely to take up the motor invitation of a timbre, in a sense co-experiencing the actions involved in a sound's production. The singing voice is particularly direct as a locus of motor resonance, as previously mentioned, though noisy signals in aggregate (including instruments) drive SMA activity among empathic people as well. This would seem to support neurophysiological studies demonstrating a robust correlation between strength of activity in pre- and primary motor cortices and empathic concern (measured by the IRI) when observing emotional behaviors of others (Carr et al., 2003; Pfeifer et al., 2008).

Egermann and McAdams (2013) found an effect of self-reported empathy on affective valence, and positively valenced non-verbal vocalizations were reported in another study to elicit a stronger motor resonance response (Warren et al., 2006). Differing from previous studies, we found instead that empathic concern covaries with *negatively* valenced ("noisy") sounds. This provides tentative evidence for our hypothesis that dispositionally empathic people would tend to exhibit greater motor susceptibility to timbral qualities that betoken heightened arousal, bodily exertion, and ecological urgency. We speculate that social- and musical empathy are closely related; empathic concern for the well-being of others is thus correlated with greater motor attunement toward others' expressions of anger, pain, danger, and other high-potency affective contexts.

V. CONCLUSION

This study used fMRI to explore the neural correlates of empathic concern in the processing of certain features of isolated timbres. We present preliminary neurophysiological evidence for the role of motor regions, especially the supplementary motor area (SMA), in the perception of "noisy" qualities of timbre by empathic individuals. We interpret a number of novel findings through the framework of embodied cognition to posit a functional link between motor systems, timbre perception, and the dispositional trait of empathic concern (as evaluated using the IRI). This is the first study to suggest that empathy modulates not only music processing, as has been explored elsewhere (Egermann & McAdams, 2013; Vuoskoski & Eerola, 2012), but also the perception of individual timbres removed from any musical context.

There are a number of limitations to the present study. Future researchers might find it fruitful to investigate the somatotopy of motor resonance in empathy-modulated timbre perception. Localization tasks for vocal production and instrument-specific actions might help disambiguate the relationship between gross motor resonance and muscle-specific co-activation in sound production and perception. Further, it remains unclear how timbre-specific effects of empathy relate (if at all) to music listening more generally.

ACKNOWLEDGMENT

We would like to thank Roger Kendall, Katy Cross, and Marita Meyer for their assistance at various stages of this study. Our research was sponsored by a UCLA Transdisciplinary Seed Grant.

REFERENCES

Aglioti, S., & Pazzaglia, M. (2010). Representing actions through their sound. *Experimental Brain Research, 206*, 141–151.

Alluri, V., *et al*. (2012). Large-scale brain networks emerge from dynamic processing of musical timbre, key, and rhythm. *Neuroimage, 59*, 3677–3689.

Beckman, C., Jenkinson, M., & Smith, S. M. (2003). General multi-level linear modelling for group analysis in fMRI. *Neuroimage*, *20*, 1052–1063.

Blumstein, D., Bryant, G. A., & Kaye, P. (2012). The sound of arousal in music is context-dependent. *Biology Letters, 8*, 744–747.

Brown, S., Ngan, E., & Liotti, M. (2008). A larynx area in the human motor cortex. *Cerebral Cortex, 18*, 837–845.

Carr, L., Iacoboni, M., Dubeau, M.-C., Mazziota, J. C., & Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas. *Proc. Nat'l Academy of Sciences*, *100*, 5497–5502.

Christov-Moore, L., & Iacoboni, M. (2016). Self-other resonance, its control and prosocial inclinations: Brain-behavior relationships. *Human Brain Mapping*, *37*, 1544–1558.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*, 113–126.

Egermann, H., & McAdams, S. (2013). Empathy and emotional contagion as a link between recognized and felt emotions in music listening. *Music Perception, 31*, 139–156.

Gallese, V. (2007). Before and below 'theory of mind': Embodied simulation and the neural correlates of social cognition. *Philosophical Transactions: Biological Sciences, 362*, 659–669.

Gazzola, V., Aziz-Zadeh, L., & Keysers, C. (2006). Empathy and the somatotopic auditory mirror system in humans." *Current Biology*, *16*, 1824–1829.

Halpern, A., Zatorre, R. J., Bouffard, M., & Johnson, J. A. (2004). Behavioral and neural correlates of perceived and imagined musical timbre. *Neuropsychologia, 42*, 1281–1292.

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1994). *Emotional contagion.* Cambridge: Cambridge Press.

Jenkinson, M., Bannister, P., Brady, J. M., & Smith, S. M. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, *17*, 825–841.

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*, *31*, 559–621.

Kaplan, J., & Iacoboni, M. (2006). Getting a grip on other minds: Mirror neurons, intention understanding, and cognitive empathy." *Social Neuroscience*, *1*, 175–183.

Kendall, R. A., & Carterette, E. C. (1993). Verbal attributes of simultaneous wind instrument timbres: I. von Bismarck adjectives." *Music Perception*, *10*, 445–467.

Lartillot, O., & Toiviainen, P. (2007) A Matlab toolbox for musical feature extraction from audio. *Proc. 10th Int'l Conf. Digital Audio Effects*, Bordeux, 237–244.

Launay, J. (2015). Musical sounds, motor resonance, and detectable agency. *Empirical Musicology Review, 10*, 30–40.

Leman, M. (2007). *Embodied music cognition and mediation technology.* Cambridge, MA: MIT Press.

Lima, C. E., *et al.* (2015). Feel the noise: Relating individual differences in auditory imagery to the structure and function of sensorimotor systems. *Cerebral Cortex, 25*, 4638–4650.
Margulis, E. H., Mlsna, L. M., Uppunda, A. K., Parrish, T. B., & Wong, P. C. M. (2009). Selective neurophysiologic responses to music in instrumentalists with different listening biographies. *Human Brain Mapping*, *30*, 267–275.
Menon, V., *et al.* (2002). "Neural correlates of timbre changes in harmonic sounds. *Neuroimage, 17*, 1742–1754.
Molnar-Szakacs, I., & Overy, K. (2006). Music and mirror neurons: From motion to 'e'motion, *Social Cognitive and Affective Neuroscience*, *1*, 235–241.
Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M., & Fried, I. (2010). Single-neuron responses in humans during execution and observation of actions. *Current Biology*, *20*, 750–756.
Nguyen, V. T., Breakspear, M., & Cunnington, R. (2014). Reciprocal interactions of the SMA and cingulate cortex sustain premovement activity for voluntary actions. *Journal of Neuroscience*, *34*, 16397–16407.
Osgood, C. E., Suci, G. J., & Tannenbaum P. H. (1957). *The Measurement of Meaning.* Urbana: University of Illinois Press.
Overy, K., and Molnar-Szakacs, I. (2009). Being together in time: Musical experience and the mirror neuron system. *Music Perception*, *26*, 489–504.
Pantev, C., Roberts, L. E., Schulz, M., Engelien, A., & Ross, B. (2001). Timbre-specific enhancements of auditory cortical representations in musicians. *Neuroreport, 12*, 169–174.
Pfeifer, J. H., Iacoboni, M., Mazziota, J. C., & Dapretto, M. (2008). Mirroring others' emotions relates to empathy and interpersonal competence in children. *Neuroimage, 39*, 2076–2085.
Platel, H., *et al*. (1997). The structural components of music perception: A functional anatomical study. *Brain*, *120*, 229–243.
Preston, S. D., & de Waal, F. B. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences*, *25*, 1–20.
Roland, P. E., Larsen, B., Lassen, N. A., & Skinhøy, E. (1980). Supplementary motor area and other cortical areas in organization of voluntary movements in man. *Journal of Neurophysiology, 43*, 118–136.
Scherer, K. R. (1995). Expression of emotion in voice and music. *Journal of Voice, 9*, 235–248.
Tsai, C.-G., Wang, L.-C., Wang, S.-F., Shau, Y.-W., & Hsiao, T.-Y. (2010). Aggressiveness of the growl-like timbre: Acoustic characteristics, musical implications, and biomechanical mechanisms. *Music Perception*, *27*, 209–221.
Vuoskoski, J., & Eerola, T. (2012). Empathy contributes to the intensity of music-induced emotions. In E. Cambouropoulos, C. Tsougras, P. Mavromatis, & K. Pastiadis (Eds.), *Proc. 12th Int'l Conf. on Music Perception and Cognition* (pp. 1112–1113). Thessaloniki, Greece.
Wallmark, Z. (2014). Appraising Timbre. PhD diss, UCLA.
———,Iacoboni, M., Deblieck, C., & Kendall, R. A. (in prep). Perceptual and neural correlates of timbral noisiness.
Warren, J. E., et al. (2006). Positive emotions preferentially engage an audio-motor 'mirror' system. *Journal of Neuroscience*, *26*, 13067–13075.
Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: Auditory-motor interactions in music perception and production. *Nature Reviews Neuroscience*, *8*, 547–558.

SIMPHONY: Studying the impact music practice has on neurodevelopment in youth

John R. Iversen[1,2], Johanna Baker[2], Dalouge Smith[3] and Terry Jernigan[2]
[1]Institute for Neural Computation, University of California San Diego, USA
[2]Institute for Neural Computation, University of California San Diego, USA
[3]San Diego Youth Symphony and Conservatory, USA

SIMPHONY is a five-year longitudinal study of the impact of music training on children's neural and cognitive development. The project is embedded within a larger neurocognitive study that addresses how environment, genetics and brain interact during development. On annual visits, children (ages 5-12) complete an extensive battery of cognitive tests, imaging of brain structure, and electrophysiology. Twenty students receiving intensive instrumental music training (in the San Diego Youth Simphony's El Sistema-inspired Community Opus Project) and 142 control students are currently enrolled. Of these, 98 (17 music) have completed a second annual measurement battery to date. We will present results relating beat perception and synchronization (BPS) ability assessed with the Beat Alignment Test (BAT; Iversen et al., 2008). All measures of BPS show age-dependent improvement, but with considerable variation for any given calendar age, which is what we seek to relate to brain structural differences. Music training influenced the consistency of improvement between year 1 and year 2, with greater mean change in scores for beat perception (% correct; $p=0.017$) and synchronization with a metronome (CV of ITI, $p=0.04$). While this may, unsurprisingly, reflect training, it may also reflect test performance motivation. Brain structural differences explain individual differences in beat perception ability, which was strongly related to the volume of motor/premotor cortex ($p<0.003$), after age effects ($p<0.001$) were accounted for. The finding that motor cortex volume predicts perceptual performance is interesting, given that the assessment of beat perception (judging the alignment of a click track superimposed on music with the music's beat) and did not explicitly involve movement.

Investigating use-dependent structural adaptations and brain network connectivity: Measuring the effects of different characteristics comprising intensive musical training on the brain

Catheryn R. Shaw[1,2,3], Robin W. Wilkins[2,3,4] and Robert A. Kraft[3,4,5]
[1]School of Music, Theatre and Dance, University of North Carolina Greensboro, USA
[2]Neuroimaging Laboratory for Complex Systems, USA
[3]Joint School of Nanoscience and Nanoengineering, University of North Carolina Greensboro, USA
[4]Gateway MRI Center, University of North Carolina Greensboro, USA
[5]School of Biomedical Engineering and Sciences, Wake Forest University Baptist Medical Center, Winston-Salem, NC, USA

Research evidence suggests that sustained and intense musical training can be a powerful tool in changing the structure of the human brain. When comparing the brain structures of trained musicians to non-musicians, research evidence reveals differences can be found in multiple regions, including the corpus callosum, premotor areas, and superior temporal gyrus. However, the characteristics that define a formally trained musician are not fully consistent from study to study. While not an exhaustive list, these variations include: different total years of intensive musical training, different definition(s) of intense musical training, and different primary instrument(s) of training. This study quantified these characteristics to help further define formally trained musicians through the use of an in-depth musical training survey. The purpose was to explore whether there are training threshold characteristics at which a significant amount of structural and functional network changes in the brain occur using fMRI (functional Magnetic Resonance Imaging). Participants were grouped according to their level of musical training: (1) professional musicians (post-secondary training), (2) school-trained musicians (at least seven years of intensive 'school-age' musical training), and (3) non-musicians (at maximum receiving only the mandatory grade school general music classes). Graph theory methods were applied to analyze network connectivity properties. Results reveal unique structural and functional changes of the human brain based on use-dependent intensive training. Through the results, evidence suggests that although receiving post-secondary education in music aids in the structural and functional change of the human brain, brain network differences can also be found between those who received seven years of intensive school-aged instrumental musical training and non-musicians. These results suggest that future network neuroimaging investigations are warranted to further our understanding of how use-dependant changes occur in the brain's structural and functional network as a result of intensive musical training.

Finding the Beat: Neural Responses to Missing Pulse Rhythms

Charles S. Wasserman[1], Jung Nyo Kim[1], Edward W. Large[1] and Erika Skoe[2]
[1] Department of Psychological Sciences, University of Connecticut, Storrs, CT USA
[2] Department of Speech, Language & Hearing Sciences, University of Connecticut, Storrs, CT USA

Many rhythm perception experiments employ simple isochronous rhythms, in which synchronous neural or behavioral responses are observed. However, responses at the stimulus frequency do not allow one to distinguish whether synchrony occurs as a response to a common input, or as the result of an emergent population oscillation that entrains at a particular frequency. It is possible to create a rhythm with no spectral energy at the pulse frequency, however, by manipulating the number of events that occur anti-phase (180°) versus in-phase (0°) with the basic rhythmic cycle. Dynamical analysis predicts neural oscillation will emerge at such a "missing" pulse frequency. Previous studies have shown that subjects tap along to complex rhythms at the missing pulse frequency – a finding that supports the prediction.

This study aimed to investigate whether the sensorimotor system, as measured by 32-channel cortical EEG, would entrain to a complex rhythm at the pulse frequency even when the complex rhythm contained no spectral power at that frequency. The experiment utilized four different rhythms of varying complexity (1 simple, 2 complex, and 1 random). Fast Fourier Transform (FFT) of the Hilbert envelope showed energy at the repetition frequency (2Hz) for the simple rhythm, but no spectral energy at the missing pulse frequency (2Hz) for the complex rhythms. EEG responses to these stimuli were recorded to look for the neural oscillations at the missing pulse frequency predicted by dynamical analysis. We report evidence of a 2Hz response in the EEG to missing pulse rhythms. These data support the theory that rhythmic synchrony occurs as the result of an emergent population oscillation that entrains at this particular frequency. We also discuss generators of the 2Hz response component.

Examining motor recruitment during incongruent audiovisual music perception: A TMS Study

Chelsea Gordon[1], Ramesh Balasubramaniam[1] and Marco Iacoboni[2]
[1]Cognitive and Information Sciences, University of California - Merced, CA, USA
[2]Dept. of Psychiatry and Biobehavioral Sciences, University of California – Los Angeles, CA, USA

Music production and music perception both involve efficient integration of visual, auditory, and motor networks. Research from multiple neuroscientific methods has uncovered a large cortical and subcortical network of brain areas responsible for musical tasks, including premotor cortex, supplementary motor area and dorsolateral prefrontal regions. To date, research has focused on how visual and auditory information is integrated, as well as how the motor system is recruited during such tasks. We examine motor recruitment in music perception when auditory and visual information are at odds with one another. Using Transcranial Magnetic Stimulation and electromyography, we measure corticospinal excitability of the cortical hand area while participants experience congruent and incongruent audio-visual musical stimuli. Stimuli were either congruent audiovisual guitar playing, incongruent audio guitar with a video of a mouth singing, or incongruent singing audio with a video of guitar playing. Results, including the interpretation of corticospinal excitability measured at various time points after stimulus onset, are discussed in the context of motor simulation and theories of sensory-motor integration.

Sleep-related plasticity in auditory evoked response after a novel instrument learning

Jimmy Tobin,*[1] Takako Fujioka,*[2]

*Department of Music., Stanford University, USA
[1]jtobin1@ccrma.stanford.edu, [2]takako@ccrma.stanford.edu

ABSTRACT

It has been shown that learning experience is consolidated after sleep and that learning some aspects of sensory stimuli such as phonetic categorization is reflected in the evoked response obtained via event-related potentials (ERPs). Active engagement in goal-oriented tasks is known to facilitate sensory learning more so than passive exposure. Musicians that participate in active, sensory learning over many years experience changes in their Auditory Evoked Responses (AERs). However, it remains unclear as to how actively learning how to play an unfamiliar instrument changes auditory processing and how such an effect changes after sleep. Here, we trained musicians to play a novel virtual instrument with a synthesized sound connected using a hand-tracking game controller for about a half hour. The EEG was recorded with participants listening to this instrument's tones as well as piano tones, immediately before and after training, and a day after the learning (termed Pre, Post1, and Post2). To make the sound of the virtual instrument novel, we used a sharp attack of a struck modal bar combined with a sustained plucked string in the instrument model. The height of each hand determined the pitch (the dominant hand) and duration separately. The range of the motion was calibrated for each participant. The auditory evoked responses showed clear P1-N1-P2 complex for both stimuli, followed by a small N2-like peak around 300ms further followed by a slow positivity. Around 450-500 ms the sleep-related enhancement of a positivity was observed to be significantly different between Post1 and Post2, only in the response to the virtual instrument tones, but not to the piano tones. Moreover, this enhancement was focused in the midline and right fronto-central electrodes. Our data demonstrate the effects of sleep on the auditory processing, possibly suggesting the involvement of the right auditory cortex related to spectral timbre information processing, and medial frontal motor-related functions even during passive listening.

I. INTRODUCTION

Studies have shown that sleep plays a key role in the consolidation of memory for motor learning, spatial learning and perceptual learning (Stickgold, 2005). When focusing on perceptual learning, different sleep periods such as rapid eye movement (REM) sleep, paradoxical sleep, and slow-wave sleep are important for promoting brain plasticity and, in the adult brain, for learning and memory (Karni et al., 1994; Maquet 2001; Stickgold et al., 2000). Importantly, post-sleep improvements in perceptual tasks can be seen after the first night of sleep, after the subsequent nights of sleep, and after between-session sleeps (Censor, Karni, & Sagi, 2006). In general, adaptive procedure is more effective in improving perceptual discrimination compared to rote practice, and increasing the number of training sessions has a limit in its benefit.

The benefits of sleep for auditory perceptual learning have also been shown in phonological recognition performance following sleep (Fenn, Nusbaum, & Margoliash, 2003) and pattern discrimination of tone sequence (Atienza et al., 2004; Atienza, Cantero, & Quiroga, 2005). The latter studies, in particular, used event-related potential (ERP) to demonstrate

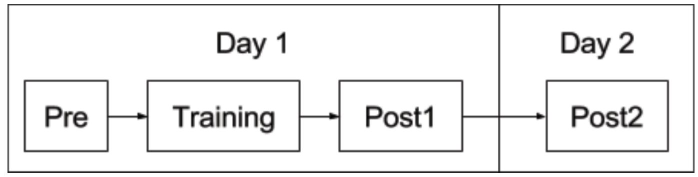


Figure 1. Schedule of training and recording sessions for each participant.

the increased mismatch negativity (MMN) component, which reflects neural process of pre-attentive perceptual discrimination, after the post-training sleep. Furthermore, perceptual learning also results in increased N1-P2 complex of the obligatory auditory ERP, or auditory evoked response (AER) (Tremblay et al., 2001; Tremblay et al., 2009), especially after post-training sleep (Alain et al., 2015). These AERs are known to show specific enhancement with musical training (Kuriki, Kanda, & Hirata, 2006; Pantev et al., 1998; Shahin et al., 2003;). Noteworthy is that the N1-P2 complex is also sensitive to music timbre (Shahin et al., 2005), in particular related to individual musical training (Pantev et al., 2001). Since training with music making activity, compared to discrimination training, was more effective in enhancing the improving behavioural discrimination of tonal patterns and accompanying MMN (Lappe et al., 2008), we hypothesized that actively learning how to play a novel instrument which produces an unfamiliar timbre and involves no physical contact with acoustic objects, would result in promoting plastic changes of auditory cortical processes for processing the timbre of this instrument, and these changes would be sensitive to the post-training sleep. We specifically designed the instrument (henceforth, 'virtual instrument') such that the past experience in playing traditional acoustic instruments do not interact with the learning. That is, the virtual instrument has to be played with arm and hand gestures, but without causing any physical excitation such as string-plucking, string-bowing, or surface-hitting. The timbre of this instrument was synthesized by combining a modal bar physical model with a plucked string physical model, to again, avoid effects of past experience in associating specific gestures with a particular instrument physical system and its sounds.

Figure 2. GameTrak controller used to play the virtual instrument

II. METHODS

A. Participants

12 young adult musicians (4 females) participated in our study. They have at least 8 years of formal musical training, training ranged from 10-30 years. The participants' dominant hand was used to keep playing the virtual instrument uniform. They gave signed informed consent for participation, and the procedure was approved by Stanford IRB board.

B. Sessions

This study took place over two days. First, the participants were fitted with a 64 channel Neuroscan Quik cap, and underwent the Pre-training EEG recording in a sound-shielded room, performed with Neuroscan SymAmpRT amplifier with 200-Hz sampling rate. This recording lasted for four eight-minute long sequences (two blocks of the virtual instrument tones, and two of piano tones, in an alternating manner with the order counter-balanced between participants). To ensure the subject was alert we asked them to press a button when they heard a white noise burst, randomly interspersed in the sequence, instead of a tone. Thereafter, the participants were engaged in training with the virtual instrument for a half hour. After the training, the EEG was recorded again immediately following the training (Post1), and the approximately 24-hours later including the overnight sleep (Post2).

C. Training

Participants were asked to play three fairly simple musical pieces (a simple one-octave arpeggio of a major chord and minor chord, Twinkle Twinkle Little Star, and Happy Birthday) shown to them as scores. Thirty minutes were given to practice these songs. Thereafter the participants were asked to perform the pieces as accurate and consistent as possible. Once the subject was able to play each piece twice with no or few mistakes, the training was finished.

D. Virtual Instrument

The virtual instrument was designed to be played with both hands wearing a GameTrak hand tracking controller (Mad Catz), which send the hand movement information to a

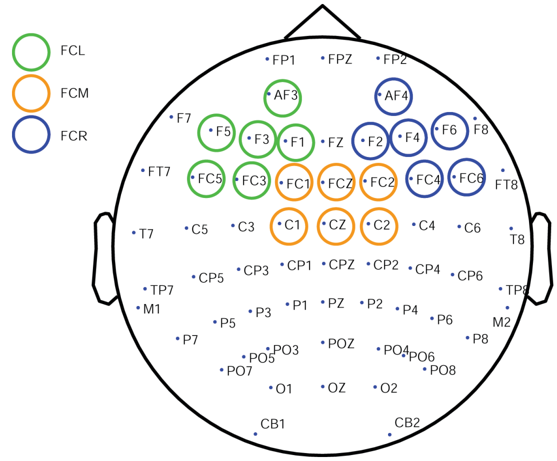


Figure 3. Electrode groupings of the 64 channel EEG cap.

computer program using ChucK (Wang, 2008) to produce musical tones with a specifically designed unfamiliar timbre (see below section E). The instrument mapped the pitch of the instrument within one-octave range (G3-G4) to the height of the player's dominant hand and the duration of the note (Half, quarter, and eighth notes at the tempo of 150 bpm) to the height of the non-dominant hand. The player would then move their hand away from them in a striking motion horizontally. The instrument would sound once the hand passed a threshold of distance from the player. All distances and notes were scaled to the players' comfortable reach in a seated position, measured at the beginning of the training.

E. Stimuli

Participants were presented with two types of tones during the EEG recording. The first is the tones of the virtual instrument, using a combination of a modal bar physical model with a plucked string physical model. The striking of the modal bar gave us a well defined note attack, while the string model made the tone novel and unfamiliar. Participants also listened to piano timbre tones as a control. Tones with both timbres

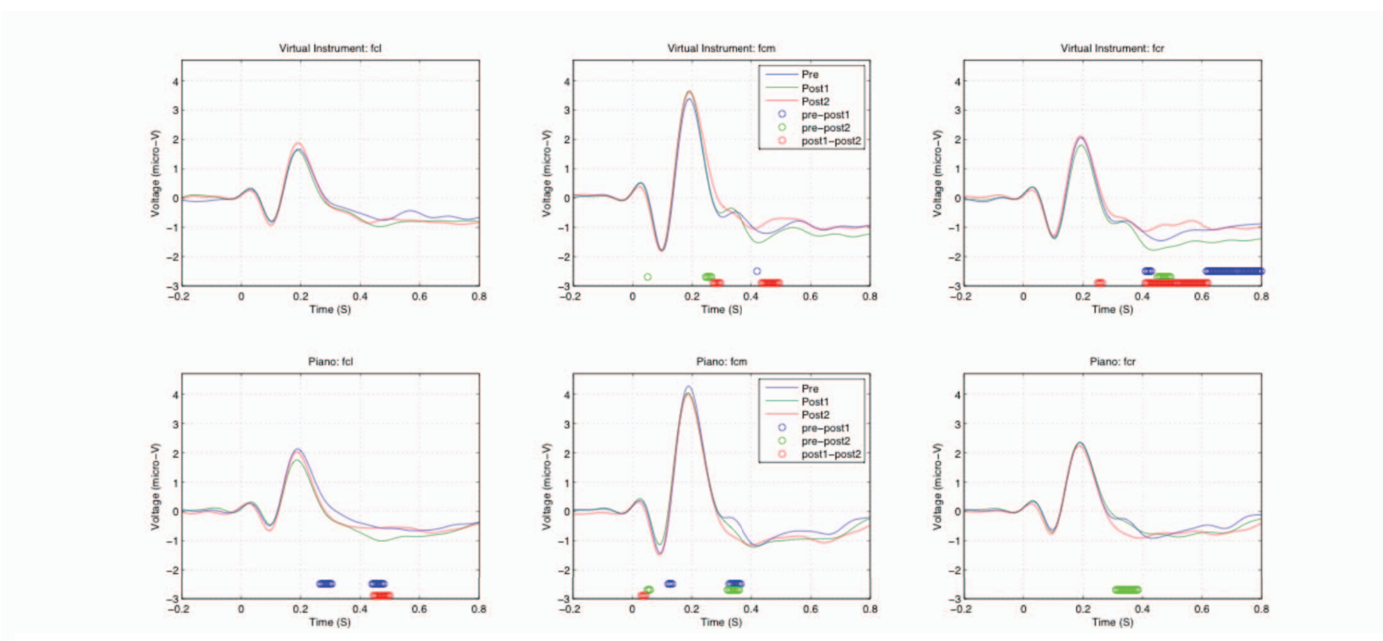


Figure 4. Auditory evoked responses to the virtual instrument and piano tones for each session across three groups of electrodes.

were edited, with 7 fundamental frequencies of C-major diatonic scale (G3 to G4), and normalized in loudness.

F. EEG Analysis

After re-referencing the EEG electrode to the common average, eye artifacts were removed using eye blinks and movements detected by horizontal and vertical electrooculogram (EOG) through source space projection in Brainstorm toolbox (Tadel et al., 2011). Afterwards the data were segmented to epochs of -0.2 to 0.8 sec to the stimulus onset, and subsequently averaged to obtain auditory evoked response for virtual instrument and piano conditions separately. The baseline correction was done using a window of -0.1 to 0 sec. Three electrode groups, namely, FrontCentral_Left (fcl), FrontCentral_Medial (fcm) and FrontCentral_Right (fcr, Figure 3) were used to inspect the waveforms and preliminary time-point-by-point pair-wise t-test comparison between any of two sessions. Based on the latency area exhibiting the session-differences, repeated measures analysis of variance (ANOVA) was used to examine the amplitude in the fcm electrode site, using the factors Timbre (virtual instrument vs. piano) and Session (Pre, Post1, Post2).

III. RESULTS

The clear auditory evoked response was obtained for all the sessions and tone types, presented in Figure 4. The significant differences between sessions Pre, Post1 and Post2 within the stimulus type, as marked in Figure 4, demonstrated that the primary effects of the sessions involved different time ranges across the virtual instrument and piano tones. In the virtual instrument condition, the effects of sleep in the fronto-central midline electrode site were evident in the two time ranges: the first was around 270ms in both Pre vs. Post2 and Post1 vs. Post2 comparisons, whereas the second range was found at 470ms in Post1 vs. Post2 comparison (Figure 4, top center panel). In contrast, for the piano condition, the sleep-related effects were obtained mainly in the 400-500ms range for Post1 vs. Post2 comparison at the left fronto-central electrode group, and in the 300-400ms range for Pre and Post2 in the midline and right electrode groups (Figure 4, bottom center and left panels). Figure 5 shows the evoked responses to the two stimulus types, their difference at each testing session, as well as the differences between two testing sessions. In the fronto-central midline electrode group, the effects of sleep were evident in three time ranges, namely around 150ms, 270ms, and 470ms. The amplitude was calculated in a 20-ms time window centered at these latencies, and submitted to three separated repeated-measures ANOVA to determine the contribution of the session and the stimulus type to the change in the signal.

A. 150ms

The ANOVA showed the significant main effect of timbre around 150ms, riding on the onset slope of P2 peak ($F(1,10) = 21.510$, $p = 0.0009$), reflecting the more negative response to the virtual instrument than piano tones. No other main effects or interactions were significant.

B. 270ms

The ANOVA showed that a negativity around 270ms, following the P2 peak around 200ms, to the virtual instrument stimulus was significantly larger from the response to the

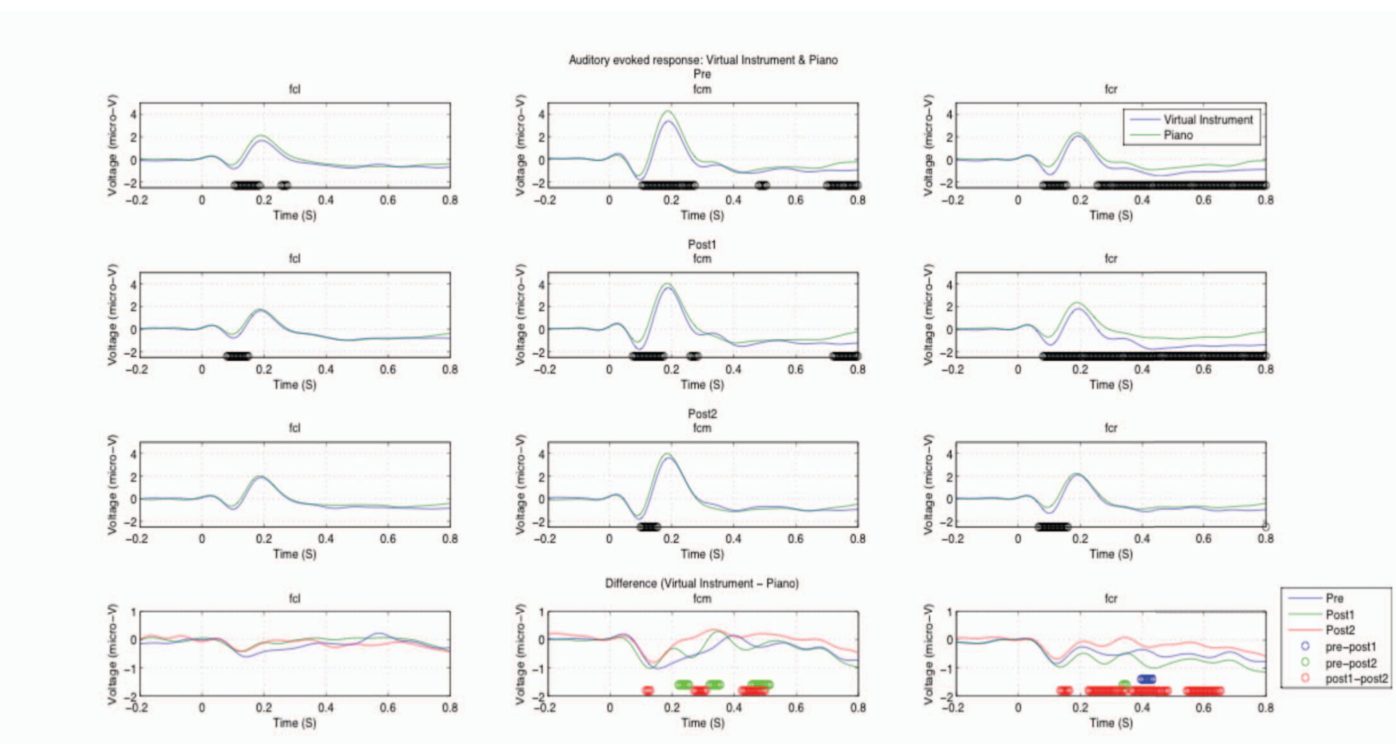


Figure 5. The evoked responses to the virtual instrument and piano stimuli, their difference at the Pre, Post1 and Post2 testing sessions, and the differences between two testing sessions. Graphs have circles underneath the curve to show time points of statistically significant difference between the stimuli (top and middle) and between sessions (bottom).

piano stimulus (F(1,10) = 7.757, p = 0.0193). No other main effects or interactions were significant.

C. 470ms

Here, the ANOVA found no main effect of Timbre or Session. Instead, there was a significant interaction between Timbre and Session (F(2,20) = 3.768, p = 0.0409). The post-hoc test showed that this interaction was caused by the significant reduction of the negativity in the Post2, compared to Post1 only for the virtual instrument (p<0.05), as the negativity to the piano stimulus did not change across the three sessions.

IV. DISCUSSION

Past experiments focusing on the specific effects of sleep found an increase in the N1-P2 complex after the first night of sleep following acoustic discrimination training. This experiment did not find any statistical increase in the P2 peak after sleep.

Instead we found differences in the N1 and N2 peaks. The former was more related to the acoustic difference between the stimuli itself, as indicated in the difference waveform between the stimuli consistently. The latter took place at the offset slope of P2, across all recording sessions of the virtual instrument. N2 is known as attention related AER component. Because participants were trained on a music making task instead of an acoustic discrimination task difference may have been caused by the brain activity related to motor and/or attentional processing even though the EEG was recorded during passive listening.

The negativity around 470ms was also found to be sensitive to the session. Interestingly, the negativity was enhanced at the Post1 recording of the virtual instrument stimulus, compared to the Pre recording, while this increment went away at the Post2 session. The nature of this peak is not really clear. However, it might be related to the N5 peak, which is usually recorded around 500ms, and found to be "related to the amount of harmonic integration required by a musical event" (Koelsch 2005). The integration of the novel timbre perception and learning might be the cause of the presence of the peak after training. The consolidation after sleep might be the reason we do not see the peak during the Post2 recording of the virtual instrument stimulus. We cannot associate the peak with N400, associated with semantic integration, because the literature shows different topography about priming effect, and perhaps different from what we observed. It is believed there is a connection between the N400 and N5 peaks, and the N5 is a measure of musical meaning. Another noteworthy aspect of the results was that these effects of training and sleep mainly involved midline and right fronto-central electrode sites, as shown in Figure 4 and 5, indicating stronger involvement of the right auditory cortex, in line with the favored timbre processing (i.e., spectral information) (Zatorre and Belin, 2001)

V. CONCLUSION

Sleep plays a role in the process of learning a new instrument for musicians. We believe that training participants in a new instrument for 30 minutes lead to an enhanced negativity around 270ms possibly related to attention and motor processing. We also saw a significant increase in the right-lateralized negativity around 500ms after training, which went away after sleep. Our findings indicate evidence of neural plasticity and the consolidation of memory during sleep for attention and timbre processing.

ACKNOWLEDGMENT

We would like to thank Professor Ge Wang (Stanford University) for allowing us to use his GameTrak controller.

REFERENCES

Alain, C., Da Zhu, K., He, Y., & Ross, B. (2015). Sleep-dependent neuroplastic changes during auditory perceptual learning. *Neurobiology of learning and memory*, *118*, 133-142.

Atienza, M., Cantero, J. L., & Quiroga, R. Q. (2005). Precise timing accounts for posttraining sleep-dependent enhancements of the auditory mismatch negativity. *NeuroImage*, *26*(2), 628-634.

Atienza, M., Cantero, J. L., & Stickgold, R. (2004). Posttraining sleep enhances automaticity in perceptual discrimination. *Journal of Cognitive Neuroscience*, *16*(1), 53-64.

Censor, N., Karni, A., & Sagi, D. (2006). A link between perceptual learning, adaptation and sleep. *Vision research*, *46*(23), 4071-4074.

Fenn, K. M., Nusbaum, H. C., & Margoliash, D. (2003). Consolidation during sleep of perceptual learning of spoken language. *Nature*, *425*(6958), 614-616.

Karni, A., Tanne, D., Rubenstein, B. S., Askenasy, J. J., & Sagi, D. (1994). Dependence on REM sleep of overnight improvement of a perceptual skill. *Science*, *265*(5172), 679-82.

Kuriki, S., Kanda, S., & Hirata, Y. (2006). Effects of musical experience on different components of MEG responses elicited by sequential piano-tones and chords. *The Journal of Neuroscience*, *26*(15), 4046-4053.

Koelsch, Stefan. "Neural substrates of processing syntax and semantics in music." *Current opinion in neurobiology* 15.2 (2005): 207-212.

Lappe, C., Herholz, S. C., Trainor, L. J., & Pantev, C. (2008). Cortical plasticity induced by short-term unimodal and multimodal musical training. *The Journal of Neuroscience*, *28*(39), 9632-9639.

Maquet, P. (2001). The role of sleep in learning and memory. *Science,294*(5544), 1048-1052

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., & Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature*,*392*(6678), 811-814.

Pantev, C., Roberts, L. E., Schulz, M., Engelien, A., & Ross, B. (2001). Timbre-specific enhancement of auditory cortical representations in musicians. *Neuroreport*, *12*(1), 169-174.

Shahin, A., Bosnyak, D. J., Trainor, L. J., & Roberts, L. E. (2003). Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *The Journal of Neuroscience*, *23*(13), 5545-5552.

Shahin, A., Roberts, L. E., Pantev, C., Trainor, L. J., & Ross, B. (2005). Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. *Neuroreport*, *16*(16), 1781-1785.

Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature*,437(7063), 1272-1278.

Stickgold, R., Whidbee, D., Schirmer, B., Patel, V. & Hobson, J. A. (2000). Visual discrimination task improvement: A multi-step process occurring during sleep. *J. Cogn. Neurosci*.12, 246–254

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., & Leahy, R. M. (2011). Brainstorm: a user-friendly application for MEG/EEG analysis. *Computational intelligence and neuroscience*, *2011*, 8.

Tremblay, K., Kraus, N., McGee, T., Ponton, C., & Otis, B. (2001). Central auditory plasticity: changes in the N1-P2 complex after speech-sound training. *Ear and hearing*, *22*(2), 79-90.

Tremblay, K. L., Shahin, A. J., Picton, T., & Ross, B. (2009). Auditory training alters the physiological detection of stimulus-specific cues in humans. *Clinical Neurophysiology*, *120*(1), 128-135.
Wang, G. (2008). The ChucK Audio Programming Language: A Strongly-timed and On-the-fly Environ/mentality. PhD Thesis, Princeton University.
Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral cortex*, *11*(10), 946-953.

A nonlinear dynamical systems approach to auditory scene analysis

Ji Chul Kim[1] and Edward W. Large[1]
[1]Department of Psychological Sciences, University of Connecticut, Storrs, Connecticut, USA

Auditory scene analysis is important both for understanding how the auditory system works and for developing computational systems capable of analyzing acoustic environments for relevant events. The auditory system parses concurrent frequencies in the acoustic signal into coherent sound events such as musical tones. Recently, nonlinear oscillatory processes were found at various stages of auditory processing, and oscillatory network models have been used to explain both neurophysiological and perceptual data. Here we show that many aspects of auditory scene analysis can be understood as signal processing and pattern formation in a dynamical system consisting of nonlinear oscillatory components tuned to distinct frequencies. We first analyze a canonical model to identify common dynamical properties of oscillatory networks that are relevant to auditory scene analysis. We focus on mode-locked synchronization and Hebbian plasticity between oscillators with harmonic frequency relationships. Next, we use the canonical model to replicate some of the key empirical findings in the field, and we compare the model to existing models and mechanisms. Simulations show that the pattern of plastic connections formed between tonotopically tuned oscillators functions similarly to a harmonic template. It provides reliable predictions for grouping and segregation of frequency components and replicates phenomena such as the formation of multiple concurrent harmonic complexes with distinct F0s and the “pop-out” of a mistuned harmonic. The canonical dynamical systems model provides a general theoretical framework which can account for many neurophysiological processes and perceptual phenomena related to auditory scene analysis.

The “musician advantage” in executive functions is more pronounced in school-age than adolescence

Katri Saarikivi[1],Vesa Putkinen[1,2], Mari Tervaniemi[1,3]
[1]Cognitive Brain Research Unit, University of Helsinki, Helsinki, Finland
[2]Department of Music, University of Jyväskylä, Jyväskylä, Finland
[3]CICERO Learning, P.O. Box 9, FI-00014, University of Helsinki, Helsinki, Finland

Several recent studies have found that adult musicians outperform non-trained peers in tasks requiring executive functions. However, this putative “musician advantage” in executive functions is still largely unexplored in children.

We investigated differences between musically trained and non-trained children (11-13y) and adolescents (15-17y) in performance in tests for executive functions and in event-related potential (ERP) responses related to involuntary attention capture (the P3a). In the ERP experiment, we presented distractor sounds in a sequence of repeating standard tones while the participants were engaged in a visual categorization task. For the neuropsychological evaluation, we administered the tests for inhibition and set-shifting from the NEPSY-II battery.

The musically trained 11-13-year-olds showed smaller P3a responses to distractor sounds compared to their non-trained peers. In the 15-17-year-olds, by contrast, there were no significant group differences in response amplitudes. However, irrespective of age, the musically trained children showed stronger habituation of the P3a response across the experiment as well as higher scores in the tests for inhibition and set-shifting.

Thus, the ERP results suggest that the musically trained children had relatively mature control over auditory attention already at age 11–13 years. By age 15–17, these skills had improved in the control children, as indexed by lack of group differences in P3a amplitude. However, they still remained below of that of the music group, as indexed by the test performance and the habituation of the P3a. The results suggest that the “musician advantage” in executive functions may be most pronounced at school age and diminishes somewhat as these functions improve with age also in nontrained children.

Melodic motif identity modulates encoding of alternating polyphonic voices

Madeline Huberth[1], Takako Fujioka[1]
[1]Department of Music, Center for Computer Research in Music and Acoustics, Stanford University, USA

Musical textures can contain one or multiple motifs that repeat in different melodic lines (voices). Previous event-related potential (ERP) studies have shown that the mismatch negativity (MMN) responses reflect encoding of melodic motifs. Also, when two voices sound simultaneously, the encoding of the upper voice is advantaged in MMN. Interestingly, our recent results showed that the upper voice advantage may extend to non-simultaneous (i.e. alternating) voices, even if voices contain the same motif. The present study is aimed at examining whether differentiation between upper and lower voices is enhanced when each voice has its own motif, which varied by pitch contour. We recorded electroencephalogram (EEG) from musicians, and compared MMN responses across voices and number of motifs present. For 20% of trials in each voice, the 5th note contained a contour-changing deviant, relative to the standard direction of that motif. Within one voice, motif entry pitch varied within one-half-octave range; the pitch separation between voices was one octave on average. Results showed that upper voice MMN was larger when each voice had its own motif, compared to when both voices had the same motif, or the upper voice was presented alone. This suggests that the presence of multiple melodic identities across voices modulates preattentive encoding of melodic lines, perhaps because melodic identity serves to differentiate streams similarly to other acoustic features.

Real or Perceived? Neural Underpinnings of Expectations in the Enjoyment of Performances

Nicole K Flaig[1], Elizabeth H Margulis[2], Carrie Kroger[2], Jung Nyo Kim[1], Edward W Large[1]
[1]University of Connecticut, USA
[2]University of Arkansas, USA

Classic accounts of musical enjoyment hold that it arises primarily from characteristics of the sound itself, but more recent studies have demonstrated that enjoyment also depends on the expectations and preconceptions that people bring to the listening experience. When wine is labeled as more expensive, it is not only perceived as tasting better, but also elicits underlying neural activity that is associated with experienced pleasantness (Plassmann, O'Doherty, Shiv, and Rangel, 2008). A more recent study has built upon this concept by examining music, finding that the experienced pleasure of music listening is also susceptible to manipulation (Margulis & Kroger, submitted). In this study, people rated their preferences for piano performances they had been told were performed by a "world-renowned concert pianist" or a "conservatory student of piano." These labels significantly affected behavioral ratings of perceived pleasantness.

In the current study, we aim to identify the neural correlates for this shift in reported enjoyment. Nonmusicians were scanned using fMRI. They heard 8 pairs of excerpts where one was performed by a student and one by a professional, and were correctly (in half the trials) or falsely (in the other half) primed about the identity of the performer. Participants used Likert-like scales to rate each performance and choose which one they preferred after each matched pair. As in the behavioral study, results suggest that enjoyment of fine music–like fine wine–can be susceptible to extrinsic information and likewise implicate specific neural correlates. Implications for the nature of aesthetic responses to music and the cognitive basis of musical pleasure are discussed.

Enhanced feature integration in musicians: Expertise modulates the additivity of the MMNm response

Niels Chr. Hansen[1,2], Andreas Højlund Nielsen[3,4], Cecilie Møller[1,5], Marcus Pearce[6] and Peter Vuust[1]

[1]Center for Music in the Brain, Aarhus University & The Royal Academy of Music Aarhus/Aalborg, Aarhus, Denmark
[2]School of Communication and Culture, Aarhus University, Aarhus, Denmark
[3]Center of Functionally Integrative Neuroscience, Aarhus University, Aarhus, Denmark
[4]Departments of Clinical Medicine and Neurology, Aarhus University, Aarhus, Denmark
[5]Department of Psychology, Aarhus University, Aarhus, Denmark
[6]Cognitive Science Research Group and Centre for Digital Music, Queen Mary University of London, London, UK

Distinguishing and integrating features of sensory input is essential to human survival and no less paramount in music perception and cognition. Yet, little is known about training-induced plasticity of neural mechanisms for auditory feature integration. This study aimed to contrast the two hypotheses that musical expertise leads to more *independent* (i.e. segregated) or more *dependent* (i.e. integrated) processing of acoustic features presented in contexts with high or low ecological validity with respect to music. To this end, mismatch negativity (MMNm) was recorded with magnetoencephalography (MEG) from 25 musicians and 25 non-musicians, exposed to sounds presented in interleaved blocks of a musical multi-feature paradigm and a classical oddball control paradigm. In addition to single deviants differing in pitch (P), intensity (I), or perceived location (L), double and triple deviants were included differing in combinations of these features (i.e. PI, IL, LP, PIL). Consistent with previous work, neural processing overlap was assessed in terms of MMNm additivity. Specifically, empirical MMNms obtained with double and triple deviants were compared to modelled MMNms corresponding to the sum of MMNms obtained with the constituent single deviants. Significantly greater MMNm under-additivity was observed in musicians compared to non-musicians, specifically for pitch-related deviants in the musical paradigm (i.e. PI, LP, and, marginally, PIL). Conversely, expertise effects were absent from the classical oddball control paradigm which used identical sounds. This novel finding supports the dependent processing hypothesis suggesting that musicians recruit overlapping neural resources facilitating more holistic representations of domain-relevant stimuli. These specialised refinements in predictive processing may enable musicians to optimally capitalise on some of the characteristic, complex variations in acoustic structure upon which music is based.

The Influence of Meter on Harmonic Syntactic Processing in Music: An ERP Study

Clemens Maidhof[1] & Rie Asano[1]
[1]Department of Systematic Musicology, University of Cologne, Cologne, Germany

Musical syntax includes tonal-harmonic structures, but also rhythm. Although several theoretical investigations stressed the importance of the interplay between tonality and rhythm (and specifically meter) for musical syntax, neuroscientific evidence is lacking. Thus, the aim of the current experiment is to investigate the influence of meter on harmonic syntactic processing in music using event-related potentials, in particular the *early right anterior negativity* (ERAN). Participants listened to musical sequences consisting of five chords of equal loudness, the final chord function being either tonally regular or irregular. The metrical importance of chords was manipulated by presenting the sequences in two blocks: each chord sequence was preceded by a one-bar percussion sequence either in a 4/4 or in a 3/4 meter. Thus, the final chord occurred either on a metrically strong (first) or weak (second) beat. To further induce a specific meter, participants had to detect rarely occurring chords with deviant timbre and judge as fast as possible whether they were on a strong or weak beat (e.g. by pressing the left button if the deviant was 'on 1' and the right button if it was 'not on 1'). To accomplish the task, participants had thus to keep the induced metrical structure over a whole chord sequence. We hypothesize that if the metrical structure influences tonal syntactic processing, the ERAN shows larger amplitudes when chord sequences are presented in a 4/4 meter than in a 3/4 meter: in the former case, an irregular final chord is on the metrically strong beat and is considered to cause a much stronger syntactic expectancy violation, compared to when an irregular final chord is on a metrically weak beat. We discuss the results of the currently ongoing analysis of the EEG and behavioral data. Results of this experiment could provide important implications for research on musical syntax, structural and temporal expectancy building, and their neurocognitive bases.

Effect of musical expertise on the coupling of auditory, motor, and limbic brain networks during music listening

Iballa Burunat[1], Valeri Tsatsishvili[2], Elvira Brattico[3,4,5], and Petri Toiviainen[1]
[1]Finnish Centre for Interdisciplinary Music Research, Department of Music, University of Jyväskylä, Finland
[2]Department of Mathematical Information Technology, University of Jyväskylä, Finland
[3]Center for Music in the Brain (MIB), Department of Clinical Medicine, Aarhus University, Aarhus, Denmark
[4]Cognitive Brain Research Unit (CBRU), Institute of Behavioral Sciences, University of Helsinki, Helsinki, Finland
[5]Advance Magnetic Imaging (AMI) Centre, Aalto University School of Science, Finland

Although music listening is known to couple auditory and motor regions and engage the mesolimbic reward system of the brain, the functional connectivity of this network has not been studied in detail as a function of musical training. By performing independent component analysis (ICA), we aimed at studying functional integration of auditory-motor networks in musicians and controls during naturalistic music listening. We performed spatial group ICA on participants' fMRI brain responses (acquired during free-listening) separately for each group in a targeted, hypothesis-driven region of interest (ROI) related to auditory and motor function. A ROI-based ICA approach improves the separation and anatomical precision of the identified spatial components. After examination of map distribution, we subsequently estimated the degree of connectivity between a priori, hypothesis-driven selected pairs of subnetworks by computing their proportion of overlap in the most stable independent components. Preliminary findings revealed (a) a trend towards high entropy and low kurtosis in the distribution of the identified components in musicians, reflecting more widespread spatial maps compared to controls; (b) strengthened connectivity between auditory and motor regions in musicians; (c) between nucleus accumbens (NAcc) and caudate (correlates of dopamine release during and prior chills while music listening); and (d) between the NAcc and the auditory cortex (a network known to predict music reward value). Results may be construed as (a) a more heterogeneous listening mode, (b) a strengthened action-perception integration during free listening, and (c,d) an increased connection between music processing and the emotion system in musicians compared to controls, likely due to musical training. These inferences speak to musicians' increased functional integration in sensorimotor and reward networks previously found to tightly interact in a musical context.

Neural changes after multimodal learning in pianists – An fMRI study

Maria Celeste Fasano[1], Enrico Glerean[2], Ben Gold[3], Dana Sheng[4], Mikko Sams[2], Josef Rauschecker[4], Peter Vuust[1] and Elvira Brattico[1]

[1]Center for Music in the Brain (MIB), Department of Clinical Medicine, Aarhus University & The Royal Academy of Music Aarhus/Aalborg, Aarhus, Denmark

[2]Brain & Mind Laboratory, Department of Neuroscience and Biomedical Engineering, Aalto University, Espoo, Finland

[3]Montreal Neurological Institute, McGill University, Montreál, Canada

[4]Laboratory of Integrative Neuroscience and Cognition, Department of Neuroscience, Georgetown University Medical Center, Washington, DC, USA

Human behavior is inherently multimodal: we use different brain systems cooperatively, combining multimodal input for a multimodal output. This is evident when pianists listen to music they know well and exhibit spontaneous activity in motor and premotor cortices. Some authors refer to the neural pathway connecting sensory and prefrontal areas via parietal and premotor cortices as the "dorsal stream" for sensorimotor integration and control, as opposed to the "ventral stream" for object identification. Here we investigated auditory-motor sequence learning in a naturalistic setting and the role of the dorsal stream of pianists as they learn complex auditory-motor sequences. We presented 10 highly skilled pianists with audio-only, video-only and audio-video recordings of a sonata by A. Scarlatti (initially unfamiliar to all pianists) in three phases. During phase 1, brain activity in all pianists was measured with functional magnetic resonance imaging while they were presented with the recordings. In phase 2, pianists learned to play the sonata by heart over the course of 4 weeks. Phase 3 was a repetition of phase 1. We examined the similarity between pianists' brain activity during stimulus presentation before and after learning by means of inter-subject correlation (ISC) analysis. We found that when presented with the video-only recording before learning the pianists' brain activity was synchronized in the visual cortex, while after learning the locus of this synchronization shifted to motor areas. For the audio-only recording, synchronization shifted from auditory to motor and visual cortices after learning. These results demonstrate crossmodal interactions during learning of complex auditory-motor sequences.

Representation of Musical Beat in Scalp Recorded EEG Responses: A Comparison of Spatial Filtering Techniques

Steven Losorelli,[*#1] Blair Kaneshiro,[*2] Jonathan Berger[*3]

[*]Center for Computer Research in Music and Acoustics, Stanford University, Stanford, CA, USA
[#]Department of Otolaryngology, Head and Neck Surgery, Stanford University School of Medicine, Stanford, CA, USA
[1]slosorelli@stanford.edu, [2]blairbo@stanford.edu, [3]brg@ccrma.stanford.edu

ABSTRACT

Background & Aims

Humans have the innate ability to perceive an underlying beat in complex musical signals. Despite the ease and speed with which our brains are able to accomplish this task, the neural mechanism underlying the cognitive processing of beat and meter are unknown, and computational extraction of beat and other acoustical features from audio remain open topics of research. Several theories propose that beat processing is a spontaneous neural response resulting from cortical entrainment of a population of neurons in response to a stimulus at the frequency of its presentation (Vialatte et al., 2010). Previous studies have shown that specific motor regions such as the basal ganglia and supplementary motor areas of the brain are implicated in beat processing (Grahn & Brett, 2007).

Steady-state evoked potentials (SS-EPs) are natural brain responses to visual or auditory information at specific frequencies. This response is used widely to study vision processing (see Norcia et al. (2015) for a recent review). In the realm of music, recent studies by Nozaradan and colleagues have shown that neural entrainment to beat and meter can manifest through a steady-state evoked potential captured in scalp-recorded electroencephalographic (EEG) data, even in the case of imagined metrical accents and in response to stimuli for which the beat frequency is not the principal frequency in the stimulus (Nozaradan et al., 2011; Nozaradan, Peretz, & Mouraux, 2012). These findings suggest that the brain responds to periodic sound patterns at specific beat frequencies as well as at imagined subharmonics related to meter. These studies took a mass-univariate approach to data analysis – that is, analyzing data from single electrodes.

Here we extend this research and take a spatial-filtering approach to analyzing EEG responses to musical stimuli with a steady beat. Spatial filtering techniques derive linear weightings of electrodes according to specific criteria. The present analysis focuses on two such techniques: Principal Components Analysis (PCA), which maximizes variance, and Reliable Components Analysis (RCA), which maximizes mutual correlation. PCA has been used in several EEG studies utilizing naturalistic music to reduce data dimensionality (Schaefer, Desain, & Farquhar, 2013; Kaneshiro et al., 2012). RCA is a recently developed technique for deriving maximally correlated components across a collection of EEG records (Dmochowski et al., 2012). RCA has been shown to derive relevant, stimulus-related information as well as plausible scalp topographies in the case of responses to audiovisual film excerpts (Dmochowski et al., 2012; Dmochowski et al, 2014), naturalistic music stimuli (Kaneshiro et al., 2014), and steady-state visual evoked potentials (Dmochowski, Greaves, & Norcia, 2015). Using these techniques, it may thus be possible to consolidate beat-related cortical activity into lower-dimensional subspaces of the data.

In the present analysis, we derive SS-EPs from spatially filtered EEG responses to naturalistic music stimuli, defined here as sensory stimuli created to be consumed and enjoyed, and not designed or created for an experimental protocol. EEG has high temporal resolution, allowing us to study neural activity on the same time scale as musical events. We use two spatial filtering techniques – PCA and RCA – and focus our analysis on low frequencies of the EEG magnitude spectra (<12 Hz), the region of interest for tempo-related information (Krumhansl, 2000). We measure the signal-to-noise ratio (SNR) at tempo-related frequencies in the EEG spectra of single PCs and RCs to assess the representation of beat-related brain activity in these single components.

Methods

For this analysis we used a set of publicly available EEG recordings of responses to naturalistic music (Kaneshiro et al., 2016). The dataset comprised dense-array (128-channel) EEG responses to four naturalistic music stimuli (Hindi pop songs) recorded from 48 participants (12 participants per song). The experimental stimuli were selected on the basis of having a steady beat and tempo throughout, as well as the presence of Western acoustical features, while avoiding participant song familiarity. The EEG responses have been preprocessed and cleaned of data artifacts. The dataset includes responses to intact stimuli, as well as to phase-scrambled controls in which the temporal coherence of the audio was disrupted.

First, beat and tempo information was computationally extracted from the audio files of the stimuli using publicly available Matlab implementations (Ellis, 2007). Next, we computed the magnitude spectra of the stimuli and visualized the results at frequencies between 0 and 12 Hz.

For PCA analysis, EEG epochs for each song were averaged across subjects, so the current focus is on phase-locked responses to the stimuli (Zanto, Snyder, & Large 2009). Next, we used singular value decomposition (SVD) to compute Principal components from the averaged data frames. RCA was computed across all unique pairs of trials for each stimulus using a publicly available Matlab implementation (Dmochowski, Greaves, & Norcia, 2015). We then used the linear weighting matrices output by each technique to project the data into the one-dimensional subspace of the first Principal or Reliable component, and computed their magnitude spectra. Scalp topographies of the components were visualized by means of a forward model (Parra et al., 2005; Dmochowski et al., 2012).

Results

The low-frequency content of the auditory stimuli used here contain many spectral peaks in the 0 to 12 Hz region. Many of these peaks are directly related to the beat frequency, while others are likely multiples of the beat frequency peaks or the result of other acoustic features present in the data.

For the responses to the intact stimuli, both spatial filtering techniques produce roughly fronto-central component topographies, as well as strong spectral peaks at the beat frequency, its harmonics, and its subharmonics. Interestingly, activity at other peak frequencies in the stimuli that were unrelated directly to the beat is reduced or absent. The EEG magnitude spectra of responses to phase-scrambled control stimuli do not show the characteristic tempo-related peaks elicited by the intact stimuli.

Acknowledgements

This research was partially supported by a grant from the Wallenberg Network Initiative: Culture, Brain, Learning. The authors thank Anthony Norcia and Jacek Dmochowski for providing helpful advice on the analyses.

REFERENCES

Dmochowski, J. P., Sajda, P., Dias, J., & Parra, L. C. (2012). Correlated Components of Ongoing EEG Point to Emotionally Laden Attention – A Possible Marker of Engagement? *Frontiers in Human Neuroscience*, *6*, 112.

Dmochowski, J. P., Bezdek, M. A., Abelson, B. P., Johnson, J. S., Schumacher, E. H., & Parra, L. C. (2014). Audience preferences are predicted by temporal reliability of neural processing. *Nature Communications*, *5*.

Dmochowski, J. P., Greaves, A. S., & Norcia, A. M. (2015). Maximally reliable spatial filtering of steady state visual evoked potentials. *NeuroImage*, *109*, 63–72.

Ellis, D. P. W. (2007). Beat Tracking by Dynamic Programming. *Journal of New Music Research*, *36*(1), 51–60.

Grahn, J. A., & Brett, M. (2007). Rhythm and beat perception in motor areas of the brain. *Journal of Cognitive Neuroscience*, *19*(5), 893–906.

Kaneshiro, B., Berger, J., Perreau Guimaraes, M., & Suppes, P. (2012). An exploration of tonal expectation using single-trial EEG classification. In *ICMPC12-ESCOM8,* 509-515.

Kaneshiro, B., Dmochowski, J. P., Norcia, A. M., & Berger, J. (2014). Toward an objective measure of listener engagement with natural music using inter-subject EEG correlation. In *ICMPC13-APSCOM5.*

Kaneshiro, B., Nguyen, D. T., Dmochowski, J. P., Norcia, A. M., & Berger, J. (2016). Naturalistic Music EEG Dataset – Hindi (NMED-H). *Stanford Digital Repository*. Retrieved from http://purl.stanford.edu/sd922db3535

Krumhansl, C. L. (2000). Rhythm and pitch in music cognition. *Psychological Bulletin*, *126*(1), 159–179.

Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottereau, B. R., & Rossion, B. (2015). The steady-state visual evoked potential in vision research: a review. *Journal of Vision*, *15*(6), 4.

Nozaradan, S., Peretz, I., Missal, M., & Mouraux, A. (2011). Tagging the neuronal entrainment to beat and meter. *The Journal of Neuroscience*, *31*(28), 10234–40.

Nozaradan, S., Peretz, I., & Mouraux, A. (2012). Selective neuronal entrainment to the beat and meter embedded in a musical rhythm. *The Journal of Neuroscience*, *32*(49).

Parra, L. C., Spence, C. D., Gerson, A. D., & Sajda, P (2005). Recipes for the linear analysis of EEG. *NeuroImage*, *28*(2), 326-341.

Schaefer, R. S., Desain, P., & Farquhar, J. (2013). Shared processing of perception and imagery of music in decomposed EEG. *NeuroImage*, *70*, 317–326.

Vialatte, F. B., Maurice, M., Dauwels, J., & Cichocki, A. (2010). Steady-state visually evoked potentials: Focus on essential paradigms and future perspectives. *Progress in Neurobiology*, *90*(4), 418-438.

Zanto, T. P., Snyder, J. S., & Large, E. W. (2006). Neural correlates of rhythmic expectancy. *Advances in Cognitive Psychology*, *2*(2-3), 221-231.

Reading the musical mind: from perception to imagery, via pattern classification of fMRI data

Tudor Popescu[1], Hannes Ruge[1], Oren Boneh[1], Fernando Bravo[1] and Martin Rohrmeier[1]
[1]Institute for Arts and Musicology, TU Dresden, Dresden, Germany

Musical imagery – the ability to "hear" music in our mind's ear in the absence of external stimuli – has long been posited as a quasi-perceptual phenomenon. This conjecture has, in recent decades, found further support in neuroimaging findings of a significant overlap in the neural resources underlying imagery and perception for the auditory – as well as other – modalities. For instance, auditory cortex activation is reliably observed during silence, if imagery is triggered by presenting music with silent gaps or silent videos strongly implying sound. Musical imagery is driven by expectations based largely on implicit musical knowledge, but while the role of expectation in music has been well studied in perception, this holds less for musical imagery. Here we used multivariate pattern analysis (MVPA), a technique that describes the extended brain spatial representations that distinguish stimuli in various conditions. We used an fMRI passive-listening paradigm that established a harmonic context via the presentation of in-key chords (cadences). Cadences in each of several categories created a strong expectation for certain end-chords, which were presented in one condition ("perception") but omitted in the other ("imagery"). An MVPA classifier was trained to distinguish between these former categories, i.e. between phrases leading to one heard end-chord or another. Pilot data so-far supports our hypothesis that the classifier trained in the perception condition generalises to the imagery condition, such that an end-chord's identity can be inferred from brain activity (reflecting imagery) acquired during the silence replacing that end-chord in its harmonic context. This result would suggest the existence of an extended network underlying both perception and imagery, one which encodes stimulus identity regardless of modality. This finding has theoretical implications about the generative model that is argued to give rise to both musical perception and musical imagery.

Time-Dependent Changes in the Neural Processing of Repetitive "Minimalist" Stimuli

Tysen Dauer[1], Audrey Ho[2], Takako Fujioka[3]
[1]Department of Music, Stanford University, USA
[2]Departments of Electrical Engineering and Computer Science, Stanford University, USA
[3]Department of Music, Center for Computer Research in Music and Acoustics, Stanford University, USA

A wide variety of musical repertoires, including 1960's American minimalist music, are built of highly repetitive musical features. Some minimalist music used short motives each repeated for various lengths of time, similar to typical auditory Event-Related Potential (ERP) paradigms. In fact, melodic pattern difference is known to be reflected in the Mismatch Negativity (MMN) component. Also, the subsequent P3a component is associated with involuntary attentional orienting towards the deviation and its novelty. We were interested in how MMN and P3a are related to the regularity of melodic block transition. We hypothesized that, compared to the previously repeated melody (i.e. the Standard stimulus), MMN and P3a would be elicited when a new melodic block starts (i.e. the Deviant stimulus). We also expected that P3a would attenuate more over the course of irregularly repeating stimuli than in regularly repeating stimuli. We created two stimulus sequences from five different four-note melodies, each starting with a unique note. In the Constant sequence, each melodic block consisted of five repetitions of a melody in a randomized order. In the Variant sequence, the number of repetitions varied from three to seven. Ten participants heard both sequences twice while EEG was recorded. Our analysis shows that the ERPs differed between the conditions in the first and final thirds of the sequences: Constant conditions show greater MMN than Variant in the beginning, but the reverse was true in the final third of the block. Interestingly, the P3a stayed larger in the Constant condition. The Variant condition also shows strong MMN throughout the sequence suggesting that pre-attentive processing remains steady through the Variant condition. Decreased MMN for the Constant condition when comparing the beginning with the end may result from the predictability of the melodic unit repetition. The attenuated P3a found in the Variant End sequence suggests lower attentional engagement.

Utilizing multisensory integration to improve psychoacoustic alarm design in the intensive care unit

Joseph J. Schlesinger[1]
[1]Vanderbilt University Medical Center, USA; email: joseph.j.schlesinger@vanderbilt.edu

Alarms in the ICU sound frequently and 85-99% of cases do not require clinical intervention. As alarm frequency increases, clinicians develop ‘alarm fatigue’ resulting in desensitization, missed alarms, and delayed responses. This is dangerous for the patient when an alarm-provoking event requires clinical intervention but is inadvertently missed. Alarm fatigue can also cause clinicians to: set alarm parameters outside effective ranges to decrease alarm occurrence, decrease alarm volumes to an inaudible level; silence frequently insignificant alarms; and be unable to distinguish alarm urgency. Since false alarm and clinically insignificant alarm rates reach 80-99%, practitioners distrust alarms, lose confidence in their significance, and manifest alarm fatigue. Yet, failure to respond to the infrequent clinically significant alarm may lead to poor patient outcomes. Fatigue from alarm amplitude and nonspecific alarms from uniform uninformative alarms is the post-monitor problem that can be addressed by understanding the psychoacoustic properties of alarms and the aural perception of clinicians.

Our experimental paradigm will determine near-threshold auditory perception of alarms, and then use clinical scenarios to determine the stimulus-response relationships for changes in auditory alarm intensity, spanning negative to positive signal-to-noise ratios (SNRs), when performing an audiovisual secondary task designed to tax attentional and decisional resources. The result will be a stimulus-response curve in dB above ambient noise.

Results show near-threshold auditory perception of alarms is around -27 decibels (dB) from background noise at 60 dB. Additionally, with visual offset of a patient monitor, there is preserved performance measured by response time and accuracy to a clinical task at *negative* SNRs compared to positive SNRs with worsening as the SNR approaches the near-threshold of hearing. Thus, clinician performance is maintained with alarms that are *softer* than background noise.

Examining Error Processing in Musical Joint Action

Anita Paas[1], Giacomo Novembre[1], Claudia Lappe[2], Catherine J. Stevens[1], Peter E. Keller[1]
[1]MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Sydney, NSW, Australia
[2]University of Münster, Münster, Germany

Errors are rare but important events that have consequences and require adaptation of action. When acting together, these consequences are often shared amongst partners. During joint action, people need to monitor their own activity, the activity of partners, and progress toward shared action goals. However, when performing the same action at the same time and with the same expected outcome, it can be difficult to tell who is responsible for the outcome; the agency of the action becomes ambiguous. Previous research has shown that people slow down their actions after making an error, and also when a partner makes an error. Electroencephalography studies have shown consistent patterns of electrical activity in the brain when an error is made, both for one's own errors and errors made by a partner. However, previous studies have not investigated how agency affects these patterns. We propose a method for studying how ambiguity of agency affects these error-related event-related potentials (ERPs) for one's own errors and errors of a partner. We expect that agency ambiguity will affect the later, positive ERPs for both own and other's errors. Music is an ideal way to investigate this as it is a social activity and musicians are familiar with playing with other musicians. Paired pianists, who can hear but not see one another, played technical exercises in unison and in octave parts. In the unison condition, agency is relatively ambiguous, enabling the comparison of behavioural and neural measures between errors made when agency is ambiguous and non-ambiguous. Preliminary behavioural results suggest that pianists adapt to errors differently when agency is ambiguous. In this study, errors are made spontaneously and naturally. Therefore, results will indicate how pianists adjust their behaviour to errors they have committed and to those made by a partner, and how agency ambiguity affects error processing in the brain.

Perceived groove and neuronal entrainment to the beat

Gabriel A. Nespoli[1] and Frank A. Russo[1]
[1]Ryerson University, Toronto, Ontario, Canada

Groove is notoriously challenging to define but generally thought of with respect to the propensity for a piece of music to induce movement (Madison, 2006). Consistent with this view, Janata, Tomic, and Haberman (2012) showed that spontaneous movement is more likely in high- vs. low-groove music. This is perhaps not surprising considering that brain networks underlying beat processing encompass several motor areas such as premotor cortex, cerebellum, and supplementary motor area (Grahn & Brett, 2007; Grahn & Rowe, 2009). From the lens of M/EEG, neural entrainment to the beat has been found to vary as a function of stimulus parameters as well as listener history (e.g., Nozaradan, Peretz, & Mouraux, 2012; Doelling & Poeppel, 2015). On the basis of these findings, we reasoned that the extent of entrainment to the beat in motor areas might serve as a useful neural marker for "feeling the groove." Participants were asked to listen to 30-second excerpts of popular songs that had been categorized as low-, mid-, or high-groove in a previous study (Janata et al., 2012). Participants rated each excerpt on the extent to which they thought the music grooved. Sixty-four channel electroencephalography (EEG) was recorded and independent components analysis (ICA) was used to identify sources of neuro-electric activity. Fourier analyses of motor sources revealed that the amount of entrainment to the beat was greater for high- and mid- versus low-groove songs. Entrainment was also found in visual areas, but was not found to be modulated by groove. These findings support the view that a feeling of groove is related to a propensity to move.

Validation And Psychometric Features Of A Music Perception Battery For Children From Five To Eleven Years Old: Evidences Of Construct Validity And M-Factor's Reliability

Hugo Cogo-Moreira

Federal University of São Paulo, Department of Psychiatry, Brazil

hugocogobr@gmail.com

ABSTRACT

At present, few psychometric instruments support the assessment of music perception in subjects under 10 years of age, limiting opportunities to probe this facet of neural development. Here we present a brief computerized assessment grounded in modern item response theory that interrogates the perception of five specific factors (i.e. meter, duration, loudness, contour, and timbre). We validated the instrument, finding strong evidence that the items within each of the five factors converge to inform a general musical perception factor, the m-factor. The m-factor offers a new universal non-verbal measure of auditory stimulus apprehension, suitable to study the underlying neurobiology of music perception, the etiology of speech and language disorders, and innate determinants of musicality.

I. INTRODUCTION

The area of psychometrics has had a huge development specially related to computational improvements and their availability to research. However, Embretson (2004, p. 8) described as still limited the application and use of such psychometric developments. She pointed out most psychological tests still were based on classical test theory. Although such statements concern psychological tests, they are relevant to music perception and music cognition.

Tests, batteries, and scales proposed to evaluate music perception skills among youth are rare. The test developed by Wing (1948) assesses the youngest population, beginning from eight years of age. Its internal consistency was 0.91 obtained via split-half method and a test-retest ranging from 0.76 to 0.88. However, split-half methods have serious limitations, being not "readily used for purpose of internal estimation of reliability" Raykov and Marcoulides (2011, p. 152). Clearly, such indices are remote from classical test theory, criticized by (Embretson, 2004) facing the development and advances proposed by item response theory and, in a more broad sense, structural equation modeling.

Even other common used tests in the adult/adolescent population as the Montreal Battery Evaluation of Amusia (MBEA), proposed by Peretz, Champod, and Hyde (2003) developed to assess amusia, Profile of Music Perception Skills (PROMS, proposed by Law and Zentner (2012)) assessing musical ability for normal or general adult populations; Clinical Assessment of Music Perception (CAMP, developed by Kang et al. (2009) evaluating music perception among adults with cochlear implants and Musical Ear Test (MET), exclusively measuring skills in melody and rhythm perception (Wallentin, Nielsen, Friis-Olivarius, Vuust, & Vuust, 2010), have used coefficient alpha as an indicator to give the test reliability. Such a reported psychometric index is the most frequently reported indicator of reliability for congeneric measures (unidimensional models) if a) they are true-score equivalent and errors of measurement are uncorrelated or b) errors are uncorrelated and the components loaded uniformly onto the common latent dimension (for major details about Cronbach's alpha assumptions, see (Raykov, 1997). All the previously cited tools have at least two dimensions, so alpha does not apply to them.

Other common "statistical" procedures intended to provide construct validity for music perception tasks is the principal component analysis which, per se, not a model-based procedure, being actually better described as a mathematical technique (Raykov & Marcoulides, 2012). Then, PCA is not a latent variable model, but a data reduction technique (Bartholomew, 2004) not ordinarily shedding much light on latent variable structure Borsboom (2006).

No one has yet evaluated the viability of subscales as indicators of specific music perception skills, casting doubt upon it. For example, under MBEA (Peretz et al., 2003) some evidence regarding subscales might be pointed out: a) melodic discrimination ability, as a subscale, may be a crucial target for treatment development and cognitive remediation in schizophrenia (Kantrowitz et al., 2014) and clear association between prosody perception and especially rhythm perception (Hausen, Torppa, Salmela, Vainio, & Sarkamo, 2013). Debate about the viability of subscales has reawakened recently (Reise, 2012a) Rodriguez, Reise, and Haviland (2015): although researchers have confirmed the different, multidimensional phenomena from psychology, psychopathology, and personality, the unit-weighted subscale scores often have ambiguous interpretations because their variance mostly reflects the general, not the specific, trait.

The aim of this study was to create and to investigate the construct validity for a battery developed to asses music perception skills across seven domains, under items' level, for young children from elementary school (from first to fifth grade). Secondarily, to evaluate its subscales viability.

II. METHOD

A. Sample size calculation

We considered at least 10 participants per observed indicator variable (i.e., the 80 items) as a rule-of-thumb for a lower bound of an adequate sample size calculation, totaling at least 800 children as suggested by Nunnally (1967). However, to accurately perform invariance testing, the value was inflated to 1000.

B. Schools selection

We chose a stratified random sampling of 14 elementary schools from a pool of Sao Paulo State's districts and cities where the last author had a previous agreement with the Department of Education to collect and conduct research.

C. Children selection

For each school, teachers selected an average of 14 children per grade (from first to fifth) returning on average 70 children per school. Teachers, nominated by the school principals, evaluated the children on the music perception test. Based on the school's enrollment list for each grade, we instructed the teachers on how to select randomly 14 per grade using www.random.org, giving parents five workdays to return informed consent about children's participation on this research. In cases of no interest to participate or no return of the informed consent, we selected another child using the same process. This random selection without any kind of inclusion/exclusion criteria guaranteed a very heterogenic and representative sampling of the music perception spectrum.

D. The conception of the music perception battery

There are 80 pair of stimulus (called as items and represented by letter E in the table 1 and supplementary material) covering seven domains: contour (13 items), timbre (12), meter (10), pitch (5), scale (15), duration (20), and loudness (5). The 80 pair of stimulus are provided in the attached supplementary material.

E. The battery application

Before presenting the 80 items, we played six opening testing stimuli to evaluate if the child fully understood the instructions: "You, [child's name], will listen to two sound sequences which are separated by a short silent moment. You should decide if these sound sequences are equal or different. Then, push the button different if the two sequences have any difference, no matter how small. Push the button equal if you believe that the two sequences of sounds are equal perfectly." For the six stimuli, the teacher helped the children explain why they were equal or different; repeating them until they felt the children really captured the instruction. Then, the 80 items were played and children gave correct or wrong answers, marked as 1 or 0 respectively. There was no time limit to push the button. The order of presentation was random (as showed in the supplementary material), to avoid local dependence.

F. Data collection and computer system Armonikos

A computer system has been developed, called Armonikos, non-Internet-dependent, which will be installed within personal computers running Java Virtual Machine (JVM). Java was chosen because it allows software to be operating system-independent, and aggregates great possibilities to extend the design to accomplish future tasks.

G. Evaluator training

Even an automatized battery may evince some difference in the way it is explained and applied. Hence, all the 14 evaluators passed through training with the last author to ensure that they gave the same instructions. The author instructed them to avoid demonstrating any facial expressions, gestures, or words that gave positive or negative signs of children's achievement. If children became demotivated, the author instructed teachers, they should use motivational phrases as "you are almost finishing, do not give up," "it is almost complete," but avoid "you are doing well," which could indicate positive or even negative attitudes in relation to children's correct or wrong answers.

III. STATISTICAL ANALYSIS

A. Fitting the baseline model (80 items)

Confirmatory factor analysis evaluated the fit of the two proposed multidimensional models underlying the 80 items correlated-factors model and a bifactor model. For major details about the bifactor model, its features, and related indices see Reise (2012a).

To evaluate the goodness of fit for both models, we used chi-square (χ^2), Confirmatory Fit Indices (CFI), the Tucker-Lewis index (TLI), and root mean square error approximation (RMSEA). The cutoff criteria used to determine the goodness of fit are described as following: chi-square with no statistical significance (> 0.05), RMSEA near or less than 0.06, and CFI and TLI near or greater than 0.9 (Hu & Bentler, 1999). CFI and TLI are penalized under complex, multidimensional models with many items per factor like ours: they tend to worsen as the number of variables in the model increases (Kenny & McCoach, 2003). We used the weighted least square using a diagonal weight matrix with standard errors and mean- and variance-adjusted (WLSMV) estimator and due to the cluster structure (children nested in schools), the standard errors and chi-square test of the model fit took into account such non-independence using the implementation proposed by (Asparouhov, 2005, 2006). The adopted statistical significance level was 0.05.

B. Reducing the items

All fourteen teachers reported that the battery was too long. The children tired around, on average, the 50th presented item, with the younger losing interest sooner. Hence, one of the main previous aim was to reduce the 80 items, selecting only the best indicators regarding items' factor loading. Under the bifactor model, we excluded some discordant items, the majority of composed items exhibiting non-statistically significant factor loadings. We decided to preserve the concordant-items regardless of their factor loadings, producing an equilibrated battery in terms of concordant and discordant items. For loudness, we did not exclude any item, because it already had few items.

C. Viability of Subscales

The following indices were used to better understand the viability of subscales which constitute children's musical perception: coefficient omega hierarchical (ωh) (McDonald, 1999; Zinbarg, Revelle, Yovel, & Li, 2005), omega subscale (ωs) reliability estimate (Reise, 2012b), and explained common variance (ECV).

IV. RESULTS

Among the 1,006 children, 69.9% are in public schools, 45% are male, and each grade (from first to fifth) had approximately 200 children. Five out 14 schools participating in this research were private.

A. Seven correlated factors underlying the 80 items

Statistical problems appeared in the first measurement model where the 80 items were distributed across the seven preconceived correlated factors, represented by ξ. We saw a linear dependency between two factors (ξscale and ξpitch) with the other five factors within the model: ξ(scale) with ξ(contour) $\rho = 0.999$ and ξ(pitch) with ξ(duration) $\rho = 0.971$. Such linear dependency is inadmissible as stated by Kline (2011) and we solved such statistical problems (formally described as a non-positive psi (ψ) defined matrix) by the exclusion of scale and pitch factors. Consequently, there were 60 remaining items distributed across five latent factor.

B. Five factors with 60 items

After the exclusion of scale and pitch factors, there were 60 items distributed across five factors. Standardized factor loading is the degree of association between each item with its underlying factor where the closer to one, the stronger (perfect) the correlation between the item and the factor). The goodness of fit for the five correlated-factor solution returns a satisfactory adjusted model, although the TLI was lower than adopted cut-off: $\chi^2(1700) = 2101.1211$, p-value <0.001; RMSEA = 0.015 (90% confidence interval [CI] = 0.013 to 0.017); CFI = 0.902 and TLI = 0.898. For bifactor returned a good model: $\chi^2(1650) = 1938.983$, p-value <0.001; RMSEA = 0.013 (90% confidence interval [CI] = 0.011 to 0.016); CFI = 0.930 and TLI = 0.925.

C. Excluding items with non-significant p-values' factor loading

To optimize the battery, a new exclusion process left only statistically significant discordant items. This reduced the number of items close to 50 as the teachers suggested. Even for the remaining 51 items, the bifactor model is still good: $\chi^2(1173) = 1441.578$, p-value <0.001; RMSEA = 0.015 (90% confidence interval [CI] = 0.012 to 0.018); CFI = 0.922 and TLI = 0.915.

D. Viability of the five subscales

If a composite were formed based on summing the 51 items, coefficient omega hierarchical (ωH) = 0.89 and EVC = 0.69, we conclude that 89% of the variance of the resulting composite could be attributable to variance of general music perception. The reliability for the five specific factors (calculated via ω(s)) controlling for that part of the reliability due to the general perception factor were very low: ω(s)Contour = 0.113, ω(s)Duration = 0.227, ω(s)Meter = 0.124, ω(s)Timbre = 0.392, and ω(s)Loudness = 0.263.

V. DISCUSSION

Our multidimensional structure underlying the original 80 or reduced 51 items led to a bifactor model fitting better than any multidimensional correlated solution like the five correlated-factor model.

Most descriptions of validity of tools for assessing music perception skills use item parceling, grouping the items or questions posed to the participants into parcels. One then treats these parcels like directly observed variables rather than treating each individual item or question as a directly observed variable. Parceling might compromise the quality of the research by hiding problems. Under seven correlate-factors, pitch and scale within the same model presented an inadmissible solution. Both domains were linearly dependent of other latent trait (for example, contour). Hence, the ability to distinguishing scale from contour among children with no extensive music training resulted in a lack of divergent validity between such factors.

Due to the sample size, modeling at items' level was viable and we observed different patterns of positive and negative factor loadings, correlating respectively to discordant and concordant stimuli, regardless of bifactor or correlated-factor solution. Such pattern has never been described, indicating that discordant and concordant items behave differently and the inverse nature of correlation related to each hypothesized factor. Such detail regarding how the patterns of factor loading behave would be muddied under parceling procedures due to combination of multiple sources of variance into one construct (Little, Rhemtulla, Gibson, & Schoemann, 2013). Item parceling occurs when two or more items are combined (summed or averaged) prior to an analysis parcels (instead of the original items) are used as the manifest indicators of latent constructs (R. Cattell, 2013; R. B. Cattell & Burdsal Jr, 1975).

Although the idea of a general music factor had been described in Law and Zentner (2012) alluding to Charles Spearman's g-factor for intelligence, formal procedures (i.e., bifactor modeling) to evaluate its existence had not previously been conducted. Through the bifactor model, we observed that the viability and reliability of MP subscales were poor. Moreover, critical evidence was observed when ωH (0.89) was compared with omega (0.95): almost all of the reliable variance in total scores (0.89/0.95 = 0.93) can be attributed to the general factor, which is assumed to reflect individual differences in music perception taken as a whole. Only 6% (0.95 – 0.89) of the reliable variance in total scores can be attributed to the multidimensionality caused by the specific domains. Only 5% is estimated to be due to random error. In other words, the m-factor is robustly reliable even though it is multidimensional construct and the specific subdomains displayed weak viability beyond the general MP factor. Such results are not surprising, especially because in other areas of child evaluation, different constructs and their related subscales has exhibited the same poor results (Jovanović, 2015; Wagner et al., 2015).

Lastly, the total information curves for the general trait showed that the 51 items are most precise when measuring children with average music perception skills. Therefore, such a battery is less accurate when assessing gifted or impaired children in this ability.

CONCLUSION

The music perception battery showed good fit indices, especially under the bifactor model, was easy to apply in the scholarly context, and captured the main general factor of music perception with good reliability.

ACKNOWLEDGMENT

We thank FAPESP for the grant 2014/06662-8.

REFERENCES

Asparouhov, T. (2005). Sampling weights in latent variable modeling. *Structural equation modeling, 12*(3), 411-434.

Asparouhov, T. (2006). General multi-level modeling with sampling weights. *Communications in Statistics—Theory and Methods, 35*(3), 439-460.

Bartholomew, D. J. (2004). *Measuring intelligence: Facts and fallacies*: Cambridge University Press.

Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika, 71*(3), 425-440. doi:10.1007/s11336-006-1447-6

Cattell, R. (2013). Validation and Intensification of the Sixteen Personality Factor Questionnairef. *Readings in Clinical Psychology*, 241.

Cattell, R. B., & Burdsal Jr, C. A. (1975). The radial parcel double factoring design: A solution to the item-vs-parcel controversy. *Multivariate behavioral research, 10*(2), 165-179.

Embretson, S. E. (2004). FOCUS ARTICLE: The Second Century of Ability Testing: Some Predictions and Speculations. *Measurement: Interdisciplinary Research and Perspectives, 2*(1), 1-32.

Hausen, M., Torppa, R., Salmela, V. R., Vainio, M., & Sarkamo, T. (2013). Music and speech prosody: a common rhythm. *Front Psychol, 4*, 566. doi:10.3389/fpsyg.2013.00566

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55.

Jovanović, V. (2015). Structural validity of the Mental Health Continuum-Short Form: The bifactor model of emotional, social and psychological well-being. *Personality and Individual Differences, 75*, 154-159.

Kang, R., Nimmons, G. L., Drennan, W., Longnion, J., Ruffin, C., Nie, K., . . . Rubinstein, J. (2009). Development and Validation of the University of Washington Clinical Assessment of Music Perception Test. *Ear and hearing, 30*(4), 411-418. doi:10.1097/AUD.0b013e3181a61bc0

Kantrowitz, J. T., Scaramello, N., Jakubovitz, A., Lehrfeld, J. M., Laukka, P., Elfenbein, H. A., . . . Javitt, D. C. (2014). Amusia and protolanguage impairments in schizophrenia. *Psychol Med, 44*(13), 2739-2748. doi:10.1017/s0033291714000373

Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural equation modeling, 10*(3), 333-351.

Kline, R. B. (2011). Measurement Models and Confirmatory Analysis *Principles and Practice of Structural Equation Modeling* (Third Edition ed.): Guilford Press.

Law, L. N. C., & Zentner, M. (2012). Assessing Musical Abilities Objectively: Construction and Validation of the Profile of Music Perception Skills. *PLoS One, 7*(12), e52508. doi:10.1371/journal.pone.0052508

Little, T. D., Rhemtulla, M., Gibson, K., & Schoemann, A. M. (2013). Why the Items versus Parcels Controversy Needn't Be One. *Psychological methods, 18*(3), 285-300. doi:10.1037/a0033266

McDonald, R. (1999). Test theory: A unified approach. *LEA, Mahwah, NJ.*

Nunnally, J. (1967). *Psychometric theory*. New York, NY: McGraw-Hill.

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders. The Montreal Battery of Evaluation of Amusia. *Ann N Y Acad Sci, 999*, 58-75.

Raykov, T. (1997). Scale reliability, Cronbach's coefficient alpha, and violations of essential tau-equivalence with fixed congeneric components. *Multivariate behavioral research, 32*(4), 329-353.

Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory*: Routledge.

Raykov, T., & Marcoulides, G. A. (2012). *An introduction to applied multivariate analysis*: Routledge.

Reise, S. P. (2012a). Invited Paper: The Rediscovery of Bifactor Measurement Models. *Multivariate behavioral research, 47*(5), 667-696. doi:10.1080/00273171.2012.715555

Reise, S. P. (2012b). Invited Paper: The Rediscovery of Bifactor Measurement Models. *Multivariate Behav Res, 47*(5), 667-696. doi:10.1080/00273171.2012.715555

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2015). Applying Bifactor Statistical Indices in the Evaluation of Psychological Measures. *J Pers Assess*, 1-15. doi:10.1080/00223891.2015.1089249

Wagner, F., Martel, M. M., Cogo-Moreira, H., Maia, C. R. M., Pan, P. M., Rohde, L. A., & Salum, G. A. (2015). Attention-deficit/hyperactivity disorder dimensionality: the reliable 'g'and the elusive's' dimensions. *European child & adolescent psychiatry*, 1-8.

Wallentin, M., Nielsen, A. H., Friis-Olivarius, M., Vuust, C., & Vuust, P. (2010). The Musical Ear Test, a new reliable test for measuring musical competence. *Learning and Individual Differences, 20*(3), 188-196.

Wing, H. (1948). *Tests of musical ability and appreciation: An investigation into the measurement, distribution, and development of musical capacity*. London, England: Cambridge University Press.

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and McDonald's ω H: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika, 70*(1), 123-133.

Musical EMDR: Music, arousal and emotional processing in EEG

Marijn Coers,[1] Daniel Myles,[2] Joyce de Bruin[3] & Rebecca S. Schaefer[4]

Leiden Institute for Brain and Cognition / Health, Medical and Neuropsychology Unit, Leiden University, The Netherlands
[1]m.coers@umail.leidenuniv.nl, [2]Daniel.b.Myles@gmail.com, [3]joyce_de_bruin@hotmail.com,
[4]r.s.schaefer@fsw.leidenuniv.nl

ABSTRACT

Eye movement desensitization and reprocessing (EMDR) is a therapeutic approach that has been suggested to be effective for treatment of post-traumatic stress disorder through bilateral presentation of various stimuli, however its mechanism is a subject of considerable debate. As the auditory bilaterally alternating EMDR stimulus is considered an unpleasant sound, we hypothesized that a similar low-arousal state may also be induced by music stimuli that include the prescribed bilateral tones, but are more pleasant to listen to (M-EMDR). We measured an emotional p300 component in the electroencephalogram (EEG) as well as heart rate under three conditions, namely the classical EMDR, musical EMDR (bilateral alternating music composed especially for this experiment, or M-EMDR), and a control condition of non-alternating tones. Preliminary results (n=8) indicate that the emotionally disturbing pictures do not seem to be processed differently with prescribed EMDR sounds as compared to either M-EMDR or the control condition. These results – if upheld with added data – suggest that the M-EMDR is equally useful as classical EMDR (or baseline stimulation). While still preliminary, these results provide further insight into the brain mechanism of auditory EMDR and its interactions with arousal, and suggest that the emotion modulation potential of music is just as applicable in clinical treatment of trauma, and likely more pleasurable.

I. INTRODUCTION

Eye movement desensitization and reprocessing (EMDR) is a therapeutic approach, which has been suggested to be effective in treating post-traumatic stress disorder (PTSD). Whilst some studies have shown an effect of EMDR treatment (Lee & Cuijpers, 2013), its mechanism is still unclear, and has been argued to be based on induction of a low-arousal state (Barrowcliff, Gray, Freeman & MacCulloch, 2004) as well as on dual task processing, leaving fewer resources to deal with the painful recollections (Schubert, Lee & Drummond, 2011). EMDR makes use of bilateral alternating stimulation in different sensory modalities (visual, auditory, haptic), which is thought to induce the low-arousal state or dual task (or both), numbing the vividness of emotionally disturbing stimuli, although it is a subject of considerable skepticism (Davidson & Parker, 2001).

Although there is some previous work using the p300 component in the EEG to assess the effectiveness of EMDR, this study did not use emotional material, and only looked at information processing after, rather than during the sound presentation, which does not reflect the way the method is offered in clinical practice (Lamprecht et al., 2004).

In the current study, we assess the effects of clinically used auditory EMDR stimulation on the emotional P300 component, as compared to a baseline of non-bilateral rhythmic stimulation, and EMDR stimulation that was created to be more musical, while also measuring physiologically measured arousal by means of the heart rate extracted from the electrocardiogram (ECG). In this way we can evaluate the effectiveness of the bilateral aspect of the EMDR stimulation, and contrast it with a more pleasant musical stimulus. We predict that classical EMDR and musical EMDR will both show an attenuated P300 component and a decreased arousal level as compared to the baseline condition of non-alternating single tones.

II. METHODS

Although data collection is ongoing, we here report results from the first 8 participants. EEG and ECG were measured while participants watched pictures of emotionally disturbing stimuli (deviants) and a checkerboard (standard), in line with Rozenkrantz & Polich (2008), from which the visual stimuli were also selected, taken from the International Affective Picture System (IAPS, Lang et al., 2008). The same images were shown under three auditory conditions, namely classic EMDR stimulation which consists of 196 Hz tones presented alternating from side to side at 1,3 Hz (78 events per minute, as prescribed by the EMDR method in Shapiro, 2001), a baseline block that is identical in pitch and presentation rate, but presented simultaneously at both sides and a musical EMDR, which consist of tones, identical to the tones presented in the two other blocks (same pitch and presentation rate), but with a variety of other sounds (e.g. synthesizers, breathing/ wind, echo's, reverbs) integrated to create a musical EMDR stimulation that can be presented in a bilateral alternating way. The experiment is presented in E-Prime, the EEG is collected using a Biosemi Active-2 system using 32 channels, the ECG is collected using a BioPac setup. The EEG was analysed in Brain Vision Analyzer, the ECG was analyzed in Matlab 2012b.

III. RESULTS

Although the results are still preliminary, no robust difference between the three conditions is seen. Although there is a strong effect of the emotional oddball in all conditions, the heart rate, heart rate variability and p300 amplitude do not differ between conditions.

IV. CONCLUSION

Our preliminary findings suggest that musical EMDR is equally useful as classical EMDR (or baseline stimulation). If emotional stimuli are indeed processed less intensely with auditory stimulation, which is not tested in the current study, the results argue for a dual-task mechanism in processing disturbing stimuli that is not dependent on the bilaterally presented stimuli. While still preliminary, these results provide further insight into the brain mechanism of auditory EMDR and its interactions with arousal, and suggest that the emotion modulation potential of music is just as applicable in clinical treatment of trauma, and likely more pleasurable.

REFERENCES

Barrowcliff, A. L., Gray, N. S., Freeman, T. C. & MacCulloch, M. J. (2004). Eye-movements reduce the vividness, emotional valence and electroderman arousal associated with negative autobiographical memories. *Journal of Forensic Psychiatry & Psychology. 15*(2), 325-345.

Davidson, P. R., & Parker, K. C. (2001). Eye movement desensitization and reprocessing (EMDR): a meta-analysis. *Journal of consulting and clinical psychology, 69*(2), 305.

Lamprecht, F., Köhnke, C., Lempa, W., Sack, M., Matzke, M., & Münte, T. F. (2004). Event-related potentials and EMDR treatment of post-traumatic stress disorder. *Neuroscience Research. 49,* 267-272.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical report A-8.

Lee, C. W., & Cuijpers, P. (2013). A meta-analysis of the contribution of eye movements in processing emotional memories. *Journal of behavior therapy and experimental psychiatry. 44*(2), 231- 239.

Rozenkrants, B., & Polich, J. (2008). Affective ERP processing in a visual oddball task: arousal, valence, and gender. *Clinical Neurophysiology, 119*(10), 2260-2265.

Schubert, S. J., Lee, C. W., & Drummond, P. D. (2011). The efficacy and psychophysiological correlates of dual-attention tasks in eye movement and desensitization and reprocessing (EMDR). *Journal of Anxiety Disorders. 25,* 1-11.

Shapiro, F., & Forrest, M. S. (2001). *EMDR: Eye movement desensitization and reprocessing.* New York: Guilford.

The Influence of Activating Versus Relaxing Music on Repetitive Finger Movement and Associated Motor Cortical Activity

Patricia Izbicki[1], Brandon Klinedinst[2], Paul Hibbing[2], Allison Brinkman[2] and Elizabeth Stegemöller[1,2]
[1]Department of Neuroscience, Iowa State University, Ames, Iowa, USA
[2]Department of Kinesiology, Iowa State University, Ames, Iowa, USA

It has been shown in healthy adults that style of music can differentially affect movement performance. The aim of this study was to examine movement performance and associated motor cortical activity while moving to different styles of music at two different rates in young healthy adults. Thirty-two participants were asked to perform an unconstrained finger flexion-extension movement in time with a tone only, and in time with two music conditions, relaxing music and activating music. Two rates were presented for each condition. A metronome click in the music conditions ensured participants were tapping at the correct rate. Finger movement was measured using a 2 mm sensor placed on the index finger, and bipolar surface electromyography (EMG) was recorded from the first dorsal interossieous and the extensor digitorum communis. Electroencephalography (EEG) signals were recorded from a montage of 64 scalp-surface electrodes during movement conditions and rest. Kinematic and kinetic data were obtained from the sensor and EMG data. Movement onsets were manually obtained. EEG signals were epoched relative to movement onset. Epochs were then collated for all trials across each tone rate and condition. A fast Fourier transform was applied. The power spectrum was normalized so total power in the spectrum was equal to 1 and then summed for each participant. To obtain the mean spectrum across participants, all epochs were averaged resulting in a chi-square distribution. To compare spectra between conditions, the mean spectrum of one condition was divided by the mean spectrum of the second condition resulting in an *F* distribution, allowing for statistical comparison of spectrum between groups by obtaining 95th percentile confidence limits from an *F* table using the total number of epochs in each group as the degrees of freedom. Any value below or above these limits designated a significant difference between spectrums. Results revealed that movement amplitude of repetitive finger movement did not differ between conditions. Statistically significant results were found for slow rate movement in the upper beta and gamma band for comparisons between both the music and move only conditions. The comparison between activating and relaxing music revealed significant differences in the upper beta and gamma band for only fast rate movement. These results suggest that music may modulate brain activity over sensorimotor areas, but not to the level to observe differences in movement performance. Perhaps differences will be revealed among populations that demonstrate impaired movement performance, such as those with Parkinson's disease.

Neural activations of metaphor use in music performance

Rebecca S. Schaefer[1], Maansi Desai[2,3], Deborah A. Barany[4] and Scott T. Grafton[4]
[1]Leiden Institute for Brain and Cognition / Health, Medical and Neuropsychology Unit, Leiden University, The Netherlands
[2]Music Department, University of California, Santa Barbara, USA
[3]Department of Neurological Surgery, University of California, San Francisco, USA
[4]Department of Psychological and Brain Sciences, University of California, Santa Barbara, USA

Metaphors are widely used in expressive music performance, yet the process of how semantic concepts can adjust motor performance is largely unknown. The music pedagogy literature describes varying types of metaphor as being useful (i.e. other actions, agents, moods or emotions, or complex scenes or situations), but also reports large individual differences in how these metaphors affect performance. Given how general the use of metaphors is in music performance, more knowledge of the mechanism of metaphor-driven movement can have implications for pedagogy as well as clinical use (i.e. movement rehabilitation). To investigate the neural activations related to metaphor use in piano performance, two pilot studies were performed with professional pianists using functional magnetic resonance imaging (fMRI) while performing series of button presses as melodies, and applying different metaphors. Pilot 1 (n=5) used 2 melodies, played in either a 'flowing' or 'playful' way, with behavioral data suggesting that in some pianists performance timing is mostly driven by the melody, whereas in others it is driven more by the metaphor. Substantial individual differences in the neural activations were seen related to metaphor use, ranging from differences in primary motor areas to visual and attention-related areas, suggesting individualized strategies. In pilot 2 (n=5), two melodies were played with eight different metaphors, using action and emotion metaphors varying in arousal and valence content. Although the results are preliminary, this more complex design reveals similar individual differences in neural activation patterns and behavioral results. Further single subject analyses will allow for a more structured interpretation of the different aspects of metaphor-based music performance. This exploratory work will provide a first look at the way expert musicians implement expressive performance.

Audio-visual mismatch negativity during musical score reading in musicians

Shu Yu Lin[1] , Takako Fujioka[1]
[1]Center for Computer Research in Music and Acoustics, Department of Music, Stanford University, Stanford, C.A., U.S.A.

Score reading during music performance or listening simultaneously requires auditory imagery for upcoming notes and integration of the actual sound. Previous studies have shown that mismatch negativity (MMN) was elicited in musicians when they detect pitch mismatch between the score and the sound. These studies used musical scores with a single line ('staff') with a single note sequence. However, musical scores usually contain more than one staff, in which multiple notes would be played together at a given time. Also, some musicians are accustomed to reading single staff music while others are used to reading multiple staff music. The present study tested whether musicians' ability to detect pitch mismatches between the score and the sound depends on their experiences. We recruited multiple-staff-musicians (e.g., composers, conductors and pianists) and single-staff-musicians (e.g., string and wind performers). Two conditions were tested; multiple-staff-condition (combining 3 G-clef staves) and single-staff-condition (1 G-clef). In both conditions, each staff contains 6 notes and they were randomly chosen between D4 and G#5, resulting in an atonal melody. One note out of 2nd to 6th notes was altered in the corresponding auditory stimulus. In the multiple-staff-condition, only one pitch was altered in one of the three staves. The score appeared 1sec before the auditory stimulus and remained on the screen. The MMN in the single-staff-condition was larger in the single-staff-musicians compared to multiple-staff-musicians. In contrast, for the multiple-staff-condition, the MMN was larger in the multiple-staff-musicians. In line with previous findings, the results demonstrate the effect of musical experience on audio-visual MMN in both groups of musicians.

Auditory-motor processing in neural oscillatory beta-band network after music training in healthy older adults

Takako Fujioka[1], Bernhard Ross[2,3]
[1]Department of Music, Center for Computer Research in Music and Acoustics, Stanford University, Stanford, CA, USA
[2]Rotman Research Institute, Baycrest Centre, Toronto, ON, Canada
[3]Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada

Timing processing for musical rhythm is important for both performers and listeners. Previous research has shown that listening to musical rhythm involves both auditory and sensorimotor systems. We have further demonstrated that such auditory-motor coordination is also reflected by modulation pattern of neural beta-band oscillations (13–30 Hz), traditionally associated with motor functions, when listening to auditory isochronous beats and subjectively imagining metric structures. Furthermore, it is known that the beta-band oscillations have significance in ageing-related movement disorders such as Parkinson's disease and stroke-induced disabilities. Here we hypothesized that a short-term keyboard training may promote plastic changes in auditory–motor functions in healthy older adults with the age of 60 years or more. They had little musical experience, and received 15 one-on-one 1-hour lessons to learn how to read music and play keyboard with both hands spanning over approximately one month. At the pre- and post- training time points, magnetoencephalography (MEG) was recorded while they were listening to the beat stimuli with three different tempi as well as finger-tapping with and without the beat stimuli. Neuromagnetic beta-band activity for listening and tapping reproduced the periodic modulation pattern observed in the previous studies. The brain areas involved were also in line with the previous findings on the effect of tempo. Comparison to the age-matched control group who were similarly naïve in music, but received no training between two measurements apart by a month are being conducted. The present study has implication for applying rhythmic auditory stimulation and music-making rehabilitation in aging population.

Mechanical Timing Enhances Sensory Processing Whereas Expressive Timing Enhances Later Cognitive Processing

Takayuki Nakata[1], Laurel J. Trainor[2]
[1]Department of Complex and Intelligent Systems, Future University Hakodate, Hakodate, Hokkaido, Japan
[2]Department of Psychology, Neuroscience and Behaviour, McMaster University, Hamilton, Ontario, Canada

Although mismatch negativity (MMN), an index of early processing of pitch information, has been found in response to pitch changes in music excerpts with mechanical timing (Brattico, Tervaniemi, Näätänen, & Peretz, 2006; Trainor, McDonald, Alain, 2002), music performers typically add timing expression by intentionally modifying isochronous beat timing beyond what is prescribed in music scores. Such timing expression is known to enhance neural processing at phrase boundaries (Istók at al, 2013). The present study examined how different patterns of timing expression in music influence different stages of processing during music listening. In the Large Interactive Virtual Environment (LIVE) lab, electrophysiological responses to both deviant notes (17%; pitch shifted up by 1/2 semitone) and non-deviant notes (83%) in a piano excerpt (Chopin's Etude in E major) were recorded from 13 participants (18-31 years of age). Excerpts featured 3 different types of timing expression–mechanical, expressive (note onsets following a professional pianist) and reversed (note to note duration in reverse order of 'expressive' timing), all with identical dynamic patterns. Results revealed that mechanically-timed music enhances early processing while expressively- and reverse-timed music facilitates conscious access to musical information. At fronto-central locations, MMN (average of 50 ms window around the negative peak for each individual) was significantly larger when listening to mechanically-timed compared to other excerpts ($p < .05$ for mechanical- compared to reverse-timed excerpts). A late positivity, analyzed as an index for cognitive processing of music information (350-500 ms window) was significant at left-frontal, central, and parietal locations for expressively- or reverse-timed music but not for mechanically-timed music.

Music-induced positive mood broadens the scope of auditory selective attention

Vesa Putkinen[1,2], Tommi Makkonen[2], Tuomas Eerola[1,3]
[1]Department of Music, University of Jyväskylä, Jyväskylä, Finland
[2]Cognitive Brain Research Unit, University of Helsinki, Helsinki, Finland
[3]Department of Music, Durham University, Durham, UK

Behavioral and neuroimaging studies indicate that happy mood reduces the selectivity of visual attention. For instance, individuals in an induced happy mood show heightened distraction from task-irrelevant visual stimuli and augmented neural responses to peripheral visual information presented outside the focus of attention.

In the current study, we investigated with event-related potentials (ERP) whether music-induced happy mood has comparable effects on selective attention in the auditory domain. Subjects listened to experimenter-selected sad, neutral or happy instrumental music and afterwards participated in a dichotic listening task while electroencephalography (EEG) was recorded. The task of the subjects was to detect target sounds presented to the attended ear and ignore the sounds presented to the unattended ear.

Task-irrelevant distractor sounds in the unattended channel elicited ERP responses related to early sound encoding (the N1) and bottom-up attention capture (the P3a) while target sounds in the attended channel elicited an ERP response related to top-down-controlled processing of task-relevant stimuli (the P3b). For the subjects in a happy mood, the N1 and P3a responses to the distractor sounds were enlarged while the P3b elicited by the target sounds was diminished. Behaviorally, these subjects showed heightened error rates on target trials following the distractor sounds.

Thus, the ERP and behavioral results suggest that the subjects in a happy mood allocated their attentional resources more equally across the task-relevant targets and the to-be-ignored sounds compared to subjects in a sad or neutral mood. Therefore, the current study extends previous research on the effects of mood on attention that has mostly been conducted in the visual domain and indicates that even unfamiliar instrumental music can broaden the scope of selective auditory attention via its effects on mood.

Auditory Deficits Compensated for by Visual Information: Evidence from Congenital Amusia

Xuejing Lu, [*#1], Hao T. Ho,[*2] Yanan Sun,[#^3] Blake W. Johnson,[#^4] William F. Thompson, [*#5]

[*]*Department of Psychology, Macquarie University, Australia*

[#]*ARC Centre of Excellence in Cognition and its Disorders, Australia*

[^]*Department of Cognitive Science, Macquarie University, Australia*

[1]xuejing.lu@mq.edu.au, [2]tam.ho@mq.edu.au, [3]yanan.sun@mq.edu.au,
[4]blake.johnson@mq.edu.au, [5]bill.thompson@mq.edu.au

ABSTRACT

While most normal hearing individuals can readily use prosodic information in spoken language to interpret the moods and feelings of conversational partners, people with congenital amusia, a musical disorder, report that they often rely more on facial expressions and gestures, a strategy that may compensate for proposed neurophysiological deficits in auditory processing. To test the extent to which amusic individuals rely on visual information in auditory processing, we measured event-related potentials (ERP) elicited by a change in pitch (up or down) between two sequential tones paired with the change in position (up or down) between two visually presented dots. The change in dot position was either congruent or incongruent with the change in pitch. Participants were asked to judge the direction of pitch change while ignoring the visual information. Our results revealed that amusic participants performed significantly worse in the incongruent condition than control participants. ERPs showed an enhanced N2-P3 response to incongruent AV pairings for control participants, but not for amusic participants. These findings indicate that amusics biased to rely on visual information when this is available because past experience has shown that information obtained from the auditory modality may not be reliable. We conclude that amusic individuals tend to discount auditory information in the context of audiovisual stimulus pairings.

I. INTRODUCTION

Congenital amusia is a disorder of musical abilities (Ayotte, Peretz, & Hyde, 2002; Foxton, Dean, Gee, Peretz, & Griffiths, 2004; Hyde & Peretz, 2004; Peretz, 2013; Peretz et al., 2002; Peretz & Hyde, 2003). Experimental findings indicate that congenital amusia can also affect aspects of speech perception, such as the recognition of linguistic prosody. However, this deficit occurs mainly when speech is stripped of semantic information: affected individuals rarely show difficulties with natural speech and other complex sounds (e.g., environmental sounds) in everyday life (Ayotte et al., 2002; Liu, Patel, Fourcin, & Stewart, 2010; Patel, Foxton, & Griffiths, 2005). One possible explanation for this discrepancy is that amusics make use of other available cues, such as semantic or visual information, to compensate for their auditory deficits. For example, Thompson et al. (2012) found that individuals with amusia show reduced sensitivity to prosody of spoken sentences conveying a happy, sad, tender or irritated emotional state. When queried, the same individuals reported that they often rely on facial expressions and gestures to interpret the moods and feelings of people with whom they interact. Thus, the reduced sensitivity to emotional prosody does not seem to pose a problem to amusic individuals in real life because they can fall back on cues delivered through other sensory modalities.

Indirect evidence for this explanation comes from a recent study by Albouy and colleagues (2015) showing that, like normal listeners, auditory perception in amusics benefits from the presence of informative visual information. In this behavioural study, amusic and control participants were asked to detect a deviant pitch in a 5-tone sequence, which was simultaneously presented with or without visual stimuli – 5 dots appearing consecutively from the left to the right along a horizontal line but at the same vertical position. Even though the visual stimuli provided no task-relevant (pitch) information about the deviant tone, their presence facilitated amusics' ability to detect the deviant. It is unclear why task-irrelevant visual information might enhance auditory processing in amusics. One possibility is that the visual cues in Albouy et al. (2015) provided an additional temporal signal that prepared the processing of the concurrent auditory stimuli, thereby enhancing detection of deviant pitches (Jones, 1976; Nickerson, 1973). Alternatively, it is possible that amusics considered visual stimuli as a useful cue when detecting deviant pitch because they tend to use visual information to compensate their auditory difficulties on the basis of their daily experience. Because amusics believe visual information is valuable, its presence may increase their confidence in decision making, which may benefit performance.

To clarify this issue, we required amusics and normal hearing controls to judge the direction (up or down) of pitch change in two consecutive tones. The acoustic pitch change stimuli were presented concurrently with visual cues: a sequence of two dots, the second of which appeared above or below the first one. The pitch change of the two tones was either congruent (second tone higher pitch and second dot higher location, or second tone lower pitch and second dot lower location) or incongruent (second tone higher pitch and second dot lower location, or vice versa) with the spatial change of the two dots. If the visual enhancement effect is entirely task independent, we predicted that amusics should show the same increment in performance for both AV conditions, because the temporal cues and visual information in the AV-congruent and AV-incongruent displays are identical. However if there is an element of task-dependence in the enhancement effect, then amusics should show a larger increment in performance in the AV-congruent task relative to the AV-incongruent task.

The use of congruent and incongruent AV pairings in the present study provided an interesting opportunity to obtain additional information about the neural mechanisms underlying visual enhancement of auditory information in amusics. A number of EEG studies have reported that multisensory stimuli that are incongruent in spatial location or temporal synchrony elicit a N2 component, a negative polarity event-related potential (ERP) component with a latency of about 200 msec after the multisensory stimulus onset (Forster & Pavone, 2008; Lindstrom, Paavilainen, Kujala, & Tervaniemi, 2012). This component is strongly associated with conflict detection at both a response level and a stimulus-representation level (Yeung, Botvinick, & Cohen, 2004), and its amplitude is typically larger following incongruent than congruent stimuli (Nieuwenhuis, Yeung, van den Wildenberg, & Ridderinkhof, 2003). We reasoned that if amusic individuals tend to ignore audiovisual conflicts because of an over reliance on visual information, then the N2 response to incongruent stimuli may be attenuated for amusic individuals but not control participants.

II. MATERIALS AND METHOD

A. Participants

Sixteen individuals with congenital amusia and 16 control participants took part in the present study. Participants were diagnosed as congenital amusics when their composite scores of the melodic subtests of the Montreal Battery of Evaluation of Amusia (MBEA; Peretz, Champod, & Hyde, 2003) were equal or less than 65 out of 90 points, that is, 72% correct (Liu et al., 2010). All participants were right-handed and had normal hearing and normal or corrected-to-normal vision. None reported any auditory, neurological, or psychiatric disorder. The amusic and control group were matched in terms of age, gender, and education (all $p > 0.05$). The study was approved by the Macquarie University Ethics Committee, and written informed consent for participation was obtained from all participants prior to testing.

B. Stimuli

The auditory stimuli comprised 800 msec tones, which had a flute timbre and were generated with the computer software Garageband (Version 6.0.4, Apple Inc., USA). The visual stimuli were white dots (60 × 60 pixels, screen resolution: 1980 × 1024 pixels) presented for 800 msec at the center of a black computer background. Each trial contained two stimuli. We refer to the first and second stimulus in a pair as the standard and target, respectively. The auditory standard could have one of the following tones in C major scale – C (261.63 Hz), D (293.66 Hz), E (329.63 Hz), F (349.23 Hz), and G (392 Hz). The pitch of the auditory target was shifted upward or downward by 3 and 4 semitones (small interval) or 8 and 9 semitones (large interval) with respect to the standard. The visual standard always appeared at the center of the computer screen, while the visual target could be shifted 300 pixels upward or downward along with the vertical line in relation to the standard. Auditory and visual stimuli were presented simultaneously, and auditory and visual target could be congruent or incongruent in terms of the direction of their shift (e.g., congruent: both upward; incongruent: auditory upward, but visual downward).

C. Procedure

Each trial began with a fixation-cross that appeared for 500-800 msec in the centre of the screen. Subsequently, standard and target stimuli (each of 800 msec duration) were presented consecutively with a jittered inter-stimulus interval (ISI) of 300-500 msec. Participants were asked to judge whether the auditory target was shifted upward or downward with respect to the standard while ignoring the visual information. The inter-trial interval (ITI) was also jittered between 300-500 msec. Participants underwent 240 trials in total, that is, 60 trials per interval (small and large) and congruency (congruent and incongruent). All trials were presented in randomized order. Participants were given a break after every 60 trials. They completed ten practice trials prior to each block. Feedback was provided in the practice but not in the actual experiment. Participants were seated approximately 50 cm from the computer screen in an electrically shielded and sound-attenuated room with dimmed light. All sounds were presented at a comfortable level via headphones.

D. EEG Recording

The EEG was recorded with a sampling rate of 1000 Hz from 32 Ag-AgCl electrodes placed according to the extended International 10-20 electrode system and mounted to a head-cap (EASYCAP GmbH, Germany). The left mastoid electrode served as reference electrode during the recording. Four additional electrodes were employed to monitor horizontal and vertical eye movements. Electrode impedances were below 5 kΩ. Using a SynAmps 2.0 RT amplifier (Compumedics Neuroscan, USA), the EEG was filtered online with an analogue band-pass filter (0.05-200 Hz).

E. EEG Processing

Offline processing of the EEG was performed in MATLAB (R2013b; Mathwork, USA) using the EEGLAB toolbox (Delorme & Makeig, 2004). The raw data was first downsampled to 500 Hz and re-referenced to the average of the left and right mastoids. It was then segmented into one-second epochs extending from 200 msec before to 800 msec after the onset of the target. Trials with incorrect responses or in which the potential exceeded ±150 µV were excluded from further data processing. The mean of each epoch was removed (Groppe, Makeig, & Kutas, 2009) before an independent component analysis (ICA) was performed using the runica algorithm. We used an ICA-based method to identify and reject trials with unusual activity (Delorme, Makeig, & Sejnowski, 2001). In addition, we employed the EEGLAB plugin ADJUST to automatically reject ocular ICs (Mognon, Jovicich, Bruzzone, & Buiatti, 2011). After IC rejection, the EEG data was low- and high-pass filtered with a Windowed Sinc FIR Filter (Widmann & Schröger, 2012) and a cut-off frequency at 30 Hz (Blackman window; filter order: 275) and 0.5 Hz (Blackman window; filter order: 2750), respectively. Epochs were subsequently averaged per participant, condition and task and baseline corrected by subtracting a pre-stimulus interval of 200 msec.

F. ERP Analysis

The main analysis focused on the comparison of the N2 components between amusic and control groups. Based on visual inspection of the grand averages and previous studies, the ERP deflection within time window of 260-380 msec after the stimulus onset was selected, encompassing the N2-P3 complex. For each participant and condition, we computed the mean amplitudes within the pre-defined time window. Furthermore, the scalp electrodes were grouped into two regions of interest (ROIs), specifically, anterior (FP1, FP2, F3, FC3, F7, FT7, F4, FC4, F8, and FT8) and posterior (O1, O2, P3, CP3, P7, TP7, P4, CP4, P8, and TP8). The mid-coronal electrodes (T7, C3, CZ, C4, and T8) were analyzed separately.

III. RESULTS

Repeated-measures ANOVA was conducted on the mean percentage correct (PC), with the between-subject factor Group (amusic and control) and the within-subject factors Interval (large and small) and Congruency (congruent and incongruent). Similar analyses were conducted on the mean amplitudes of ERPs with an additional factor of ROIs (anterior and posterior). This factor was excluded from the analysis on the mid-coronal electrodes.

A. Behavioural Results

We found a significant interaction between the factors Congruency and Group, $F(1, 30) = 10.24$, $p < 0.01$, $\eta_p^2 = 0.25$. Examination of this interaction revealed that the performance of amusics was significantly worse than control participants in both small (amusics: M = 0.72, SE = 0.02; controls: M = 0.86, SE = 0.03) and larger interval conditions (amusics: M =0.92, SE = 0.01; controls: M = 0.96, SE = 0.02), $F(1, 30) = 12.63$, $p = 0.001$, $\eta_p^2 = 0.30$, when the visual change in direction was incongruent with the pitch change. However, when visual and auditory information was congruent, amusics (small: M = 0.88, SE = 0.02; large: M = 0.97, SE = 0.01) performed just as well as controls (small: M = 0.93, SE = 0.02; large: M = 0.98, SE = 0.01), $F(1, 30) = 3.64$, $p = 0.07$, $\eta_p^2 = 0.11$.

B. ERP Results

We focus on the results from the ROI analysis, as the analysis of the mid-coronal electrodes yielded largely similar results. For the selected time window, that is the N2-P3

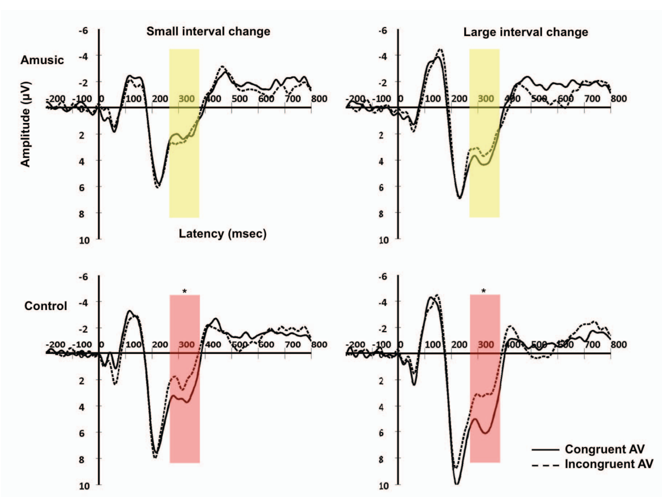


Figure 1. Grand-averaged ERPs of amusic and control groups at electrode CZ in response to auditory stimuli with congruent (solid line) and incongruent (dash line) visual stimuli. A significant congruency effect is highlighted with red bar, and a non-significant congruency effect is highlighted with yellow bar.

complex, the statistical results yielded a marginally significant interaction between Congruency and Group, $F(1, 30) = 3.89$, $p = 0.06$, $\eta_p^2 = 0.12$. It should be noted that this interaction was significant in the mid-coronal electrodes, $F(1, 30) = 6.25$, $p < 0.05$, $\eta_p^2 = 0.17$, as shown in Figure 1. For the control group, the negativity of the N2-P3 complex was larger for incongruent trials (M = 2.03 μV, SE = 0.51 μV) than for congruent trials (M = 3.01 μV, SE = 0.40 μV), $F(1, 30) = 11.68$, $p < 0.01$, $\eta_p^2 = 0.28$. For amusic group, the enhanced negativity of the N2-P3 complex for incongruent trials was not found (incongruent: M = 1.81 μV, SE = 0.51 μV; congruent: M = 1.99 μV, SE = 0.40 μV), $F(1, 30) = 0.40$, $p = 0.53$, $\eta_p^2 = 0.01$.

IV. DISCUSSION

Congenital amusia is characterized by a variety of deficits in musical perception. The present study tested the hypothesis that these auditory modality processing deficits result in a tendency in amusic individuals to rely on visual information in contexts like interpersonal conversations, when both auditory and visual cues to emotional prosody are simultaneously available.

Indeed, our data revealed that amusics' ability to identify the pitch contour was comparable to controls' when congruent visual information was supplied. However, amusics' performance deteriorated significantly when the visual information was incongruent with the auditory information. It suggests that amusic participants tend to be misled by the visual stimuli even when visual information is task-irrelevant. In other words, amusics make use of available visual cues in order to compensate for their difficulties in pitch perception. The tendency by amusics to use visual cues is compatible with the "optimal-integration hypothesis," which suggests that perceivers are more likely to rely on one modality over the other depending on how reliable the information is (Ernst & Bulthoff, 2004; Ernst & Banks, 2002). For amusic individuals, their pitch impairment means that auditory information is not reliable. Therefore, amusic individuals tend to use contextual or facial cues to boost their auditory perception (Albouy et al., 2015; Thompson, Marin, & Stewart, 2012). Regarding to ERPs results, control participants exhibited an increase in the amplitude of N2-P3 complex to incongruent relative to congruent AV pairings, suggesting a processing of conflict detection even when the visual information was task-irrelevant (Forster & Pavone, 2008; Lindstrom et al., 2012; Nieuwenhuis et al., 2003; Yeung et al., 2004). By contrast, amusics failed to show a conflict response in this task, as reflected by an absence of N2 effect. The findings in line with our behavioural data, therefore, support our idea that amusics focus on visual information to complete the auditory task thereby ignoring AV conflicts.

V. CONCLUSION

In summary, the present study is the first ERP study showing that individuals with congenital amusia rely heavily on unattended visual information when doing auditory task due to their deficits in auditory processing, providing the theoretical basis for using visual information to improve amusics' auditory perception.

REFERENCES

Albouy, P., Leveque, Y., Hyde, K. L., Bouchet, P., Tillmann, B., & Caclin, A. (2015). Boosting pitch encoding with audiovisual interactions in congenital amusia. *Neuropsychologia, 67*, 111-120.

Ayotte, J, Peretz, I., & Hyde, K. L. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain, 125*(2), 238-251.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods, 134*(1), 9-21.

Delorme, A., Makeig, S., & Sejnowski, T. (2001). *Automatic artifact rejection for EEG data using high-order statistics and independent component analysis*. Paper presented at the The third international ICA conference.

Ernst, M. O., & Bulthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences, 8*(4), 162-169.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*(24), 429-433.

Forster, B., & Pavone, E. F. (2008). Electrophysiological correlates of crossmodal visual distractor congruency effects: Evidence for response conflict. *Cognitive, Affective, & Behavioral Neuroscience, 8*(1), 65-73.

Foxton, J. M., Dean, J. L., Gee, R., Peretz, I., & Griffiths, T. D. (2004). Characterization of deficits in pitch perception underlying 'tone deafness'. *Brain, 127*, 801-810.

Groppe, D. M., Makeig, S., & Kutas, M. (2009). Identifying reliable independent components via split-half comparisons. *NeuroImage, 45*(4), 1199-1211.

Hyde, K. L., & Peretz, I. (2004). Brains that are out of tune but in time. *Psychological Science, 15*(5), 356-360.

Jones, M. R. (1976). Levels of structure in the reconstruction of temporal and spatial serial patterns. *Journal of Experimental Psychology, 2*(4), 475-488.

Lindstrom, R., Paavilainen, P., Kujala, T., & Tervaniemi, M. (2012). Processing of audiovisual associations in the human brain: dependency on expectations and rule complexity. *Frontiers in Psychology, 3*, 159.

Liu, F., Patel, A. D., Fourcin, A., & Stewart, L. (2010). Intonation processing in congenital amusia: Discrimination, identification and imitation. *Brain, 133*(6), 1682-1693.

Mognon, A., Jovicich, J., Bruzzone, L., & Buiatti, M. (2011). ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology, 48*(2), 229-240.

Nickerson, R. S. (1973). Intersensory facilitation of reaction time: energy summation or preparation enhancement? *Psychological Review, 80*(6), 489-509.

Nieuwenhuis, S., Yeung, N., van den Wildenberg, W., & Ridderinkhof, K. R. (2003). Electrophysiological correlates of anterior cingulate function in a go/no-go task: Effects of response conflict and trial type frequency. *Cognitive, Affective, & Behavioral Neuroscience, 3*(1), 17-26.

Patel, A. D., Foxton, J. M., & Griffiths, T. D. (2005). Musically tone-deaf individuals have difficulty discriminating intonation contours extracted from speech. *Brain and Cognition, 59*(3), 310-313.

Peretz, I. (2013). The biological foundations of music: Insights from congenital amusia. In D. Deustch (Ed.), *The psychology of music* (3rd Ed., pp. 551-564). San Diego: Academic Press.

Peretz, I., Ayotte, J., Zatorre, R. J., Mehler, J., Ahad, P., Penhune, V. B., & Jutras, B. (2002). Congenital amusia: A disorder of fine-grained pitch discrimination. *Neuron, 33*(2), 185-191.

Peretz, I., Champod, A. S., & Hyde, K. L. (2003). Varieties of musical disorders: The Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences, 999*(1), 58-75.

Peretz, I, & Hyde, K. L. (2003). What is specific to music processing? Insights from congenital amusia. *Trends in Cognitive Sciences, 7*(8), 362-367.

Thompson, W. F., Marin, M. M., & Stewart, L. (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proceedings of the National Academy of Sciences, 109*(46), 19027-19032.

Widmann, A., & Schröger, E. (2012). Filter effects and filter artifacts in the analysis of electrophysiological data. *Frontiers in Psychology, 3*, 233.

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review, 111*(4), 931-959.

The association of music aptitude and personality in 10- to 12-year-old children and adults

Franziska Degé[1] and Gudrun Schwarzer[1]
[1]Department of Developmental Psychology, Justus-Liebig-University, Giessen, Germany

There is empirical evidence of an association between music lessons and personality in children and adults. For older children and adults, personality (openness, conscientiousness) predicts who takes music lessons and for how long (Corrigall et al., 2013). For younger children, parent's personality (openness) predicts duration of music training (Corrigall & Schellenberg, 2015). Although such associations between music lessons and personality have been found, relations between music aptitude and personality has not been addressed. Hence, our aims were to test whether personality is associated with music aptitude as well as with music lessons, whether these associations show a similar pattern, and whether results were comparable in 10- to 12-year-old children and adults. We tested 42 (14 male) children (age: M = 10.91 years, SD = 0.76 years) and 31 (12 male) adults (age: M = 23.77 years, SD = 3.96 years). We measured as control variables IQ (CFT 20R) and socioeconomic status and as criterion measures music aptitude (AMMA), music lessons, and personality (BFI). For children and adults no significant correlations between control variables and criterion variables were found (ps > .10). Therefore, control variables were not included in main analyses. The child data showed a significant association between music aptitude and agreeableness (r = .34, p = .03). Furthermore, music lessons were significantly related to openness (r = .33, p = .04). The adult data revealed a significant association between music aptitude and openness (r = .35, p = .05) as well as a significant relation between music lessons and openness (r = .36, p = .04). We conclude that both music aptitude and music lessons are associated with personality. However, children and adults differ in personality factors that are linked to music aptitude, but not the ones linked with music lessons. Openness seems to play a crucial role in associations between personality and music lessons as well as music aptitude.

MUSEBAQ: A psychometrically robust questionnaire for capturing the many voices of music engagement

TanChyuan Chin[1], Nikki S. Rickard[1,2], Eduardo Coutinho[3] and Klaus R. Scherer[4]
[1]Graduate School of Education, University of Melbourne, Australia
[2]School of Psychological sciences, Monash University, Australia
[3]Department of Music, University of Liverpool, UK
[4]Department of Psychology, University of Geneva, Switzerland

Engagement with music is complex, influenced by music training and capacity, affective style, music preferences, and motivations. Methods to assess music engagement have included auditory tests, psychometric questionnaires and experience sampling methods. Limitations of previous research however, include: an absence of a comprehensive measure that assesses multiple aspects of music engagement; limited evidence of psychometric properties; and a bias towards equating ‘engagement’ with formal music training, overlooking other forms of strong music engagement. In this paper, we describe a newly developed and psychometrically tested modular instrument which assesses a diverse set of music engagement constructs. The MUSEBAQ can be administered in full, or by module as relevant for specific purposes. In 3 separate studies, evidence was obtained from over 3000 adults (aged 18-87; 40% males) for its structure, validity and reliability. Module 1 (Musicianship) provides a brief assessment of formal and informal music knowledge and practice. Module 2 (Musical Capacity) measures emotional music sensitivity (α=.90), listening sophistication (α=.76), indifference to music (α=.59), music memory and imagery (α =.81) and personal commitment to music (α=.80). Module 3 (Music Preferences) classifies preferences into six broad genres - rock or metal, classical, pop or easy listening, jazz, blues, country or folk, rap or hip/hop, dance or electronica. Online administration uses adaptive reasoning to selectively expand sub-genres, while minimizing time demands. Module 4 (Reasons for Music Use) assesses seven motivations for using music; musical transcendence (α=.90), emotion regulation (α=.94), social, music identity and expression (α=.90), background (α=.80), attention regulation (α=.69), and physical (α=.71). The comprehensiveness, yet flexibility, of the MUSEBAQ makes it an ideal questionnaire to use in research requiring a robust measure of music engagement.

Personality and musical preference using crowd-sourced excerpt-selection

Emily Carlson[1], Pasi Saari[1], Birgitta Burger[1] and Petri Toiviainen[1]
[1]Department of Music, University of Jyväskylä, Finland

Music preference has been related to individual differences like social identity, cognitive style, and personality, but preference can be difficult to quantify. Self-report measures may be too presumptive of shared genre definitions between listeners, while listener-ratings of expert-selected music may fail to reflect typical listeners' genre-boundaries. The current study aims to address this by using a crowd-tagging to select music for studying preference. For the current study, 2407 tracks were collected and subsampled from the Last.fm crowd-tagging service and the EchoNest platform based on attributes such as genre, tempo, and danceability. The set was further subsampled according to tempo estimates and metadata from EchoNest, resulting in 48 excerpts from 12 genres. Participants (n=210) heard and rated the excerpts, rated each genre using the Short Test of Music Preferences (STOMP), and completed the Ten-Item Personality Index (TIPI). Mean ratings correlated significantly with STOMP scores (r =.37-.83, p<.001), suggesting that crowd-sourced genre ratings can provide a fairly reliable link between perception and genre-labels. PCA of the ratings revealed four musical components: 'Danceable,' 'Jazzy.' 'Hard,' and 'Rebellious.' Component scores correlated modestly but significantly with TIPI scores (r =-.14-.20, p <.05). Openness related positively to Jazzy scores but negatively to Hard scores, linking Openness to liking of complexity. Conscientiousness related negatively to Jazzy scores, suggesting easy-going listeners more readily enjoy improvisational styles. Extraversion related negatively to Hard scores, suggesting extroverts may prefer more positive valences. Agreeableness related negatively to Rebellious scores, in line with agreeable peoples' tendency towards cooperation. These results support and expand previous findings linking personality and music preference, and provide support for a novel method of using crowd-tagging in the study of music preference.

Demographic, cognitive, and personality predictors of musicality

Matthew R. Labrecque[1] and Kathleen A. Corrigall[1]
[1]Department of Psychology, MacEwan University, Edmonton, Alberta, Canada

What kinds of people become musical? Research has begun to examine which individual difference variables predict aspects of musical expertise, such as duration of music training. However, playing a musical instrument is not the only way to exhibit musicality. The recently developed Goldsmiths Musical Sophistication Index (Gold-MSI) suggests that musicality can be conceptualized more broadly and comprises 1) active engagement (e.g., attending concerts), 2) perceptual abilities, 3) formal music training, 4) singing ability, and 5) emotional responses to music. In the current study, we compared predictors of duration of playing a musical instrument regularly with predictors of musical sophistication. We also tested previously unexamined predictor variables such as creativity and personality facets (six facets for each of the Big Five domains). We administered tests of personality (IPIP-NEO Short Form; Grit-Short Form), nonverbal IQ (modified Raven's Matrices), creativity (Guilford's Alternative Uses), and musical sophistication (Gold-MSI) to 209 undergraduate students. Musical background and demographic information (e.g., age, gender, parents' education, involvement in non-musical extracurricular activities) were also collected. Predictor variables that were significantly associated with each outcome variable were entered into multiple regression analyses. Results revealed that the predictor variables explained 19.1% of the variance in duration of playing music regularly and 32.9% of the variance in musical sophistication. Notable findings were that creativity was the strongest predictor of duration of playing music regularly, and that neither cognitive measure (non-verbal IQ, creativity) predicted musical sophistication. Our findings to date suggest that there are both overlapping as well as unique predictors of duration of playing music regularly compared to musical sophistication.

Investigating the role of parents in the perfectionistic tendencies of university music students

Madaleen Botha[1], Clorinda Panebianco-Warrens[1]
[1] University of Pretoria, Pretoria, South Africa

Perfectionism is often associated with the positive characteristics of motivation, effort and achievement. However, perfectionism encouraged by the views of others may result in increased levels of anxiety and ultimately lead to maladaptive perfectionist tendencies. Research has identified parents as important in the origin of perfectionism in children. The aim of the study is to explore the extent to which parental factors contribute to the experience of perfectionism in South African university music students. A total of 93 music students participated in the quantitative study by completing the Frost Multidimensional Perfectionism Scale (FMPS).The results showed significant differences with both the BA (Music) students and African language group scoring significantly higher than the BMus students and other languages groups in the dimensions parental expectations and parental criticism. The study provides valuable insight into the perfectionistic trends of South African undergraduate music students, with particular emphasis on the parental dimensions of perfectionism.

Can personality traits predict musical preferences? A meta-analysis

Thomas Schäfer[1]
[1]Department of Psychology, Chemnitz University of Technology, Chemnitz, Germany

Background
How can inter-individual differences in the preference for musical styles/genres be predicted? Age, gender, personality traits, social and socio-cultural variables have been discussed as the most promising variables for making good predictions. Above all, scholars have usually ascribed the largest part of the variance to inter-individual differences in personality traits. Consequently, there has been a large series of empirical studies on the relationship between personality traits and musical style/genre preferences.

Aims
Despite the large number of studies, the impression remains, however, that empirical findings are not quite consistent and that effect sizes are rather small. To examine the issue more thoroughly, the results of previous studies were collected and meta-analyzed in order to obtain reliable estimates of the effects.

Methodology
The Big 5 personality traits were most often used in previous studies and were therefore chosen as the most appropriate categorization of personality traits in the meta-analysis. The Short Test of Music Preference (STOMP) with its five dimensions was most often used as a categorization of musical style/genre preferences and therefore used in the meta-analysis. Hence, studies were included in the analysis when they had investigated the relationship between at least one of the Big 5 personality traits and at least one of the five dimensions of the STOMP. Thus, in total, there were 25 sub-analyses.

Results
All weighted averaged correlation coefficients were very small; most of them near zero. Only four of the 25 coefficients were at least r = .10. The largest effects were observed for the openness-to-experience personality trait that exhibited small correlations (between .12 and .22) with three musical styles.

Conclusion
Thus, personality traits cannot substantially account for the inter-individual variance in musical preferences. Musical functions are discussed as an alternative explanation for that variance.

‘Transmodal’ translation in ‘teaching-and-learning’ music

Annika Falthin[1]
[1]Department of Music Education, Royal College of Music, Stockholm, Sweden & University College of Music Education, Stockholm, Sweden

Based on a qualitative empirical study about affordances and meaning-making processes where students in Compulsory junior high school learn to play music together, this paper presents how multimodal instructions create embodied musical cognition. The data consist of 13 observed music lessons in two 8th grade classes, video recorded in the course of one term. In order to explore multimodal aspects of sign making and interaction in teaching and learning excerpts from teacher and students’ singing and playing music on different occasions were transcribed into scores in which musical notation together with other graphic signs and written descriptions represent the events. The scores visualise multimodal aspects of musical interaction, which made a 'fine grained' analysis of meaning-making processes possible. As the teacher rarely demonstrated the parts on the instruments the students’ interpretation process took place with the aid of the teacher’s graphic, bodily and verbal combinations of signs. In the repertoires of expressive signs cohesion was achieved largely through recognition represented by different modes. Especially conspicuous was how instructions that were linked to the beat of the music were most important for understanding, compared to signs off the beat. Significant was also how the teacher used the voice to make signs which had the effect of producing kinaesthetic experiences. The students in turn changed the sound into movements that facilitated their playing. Learning could be observed in the way in which the movements of the students became increasingly embodied during practice. Learning also revealed itself by how, in keeping time, the students made ‘transmodal’ translations of the teacher's multimodal signs. The findings exemplify the human ability to translate corresponding signs and the role of pulse and rhythm to make cohesion. In addition the findings highlight the significance of being aware of these affordances in music education.

Young children's singing modes

Eugenia Costa-Giomi[1], Lucia Benetti[1], Molly Marek[2]
[1]School of Music, The Ohio State University, Columbus, OH, USA
[2]College of Education, The University of Texas, Austin, TX, USA

Singing plays an important role in the upbringing of infants as it modulates arousal (Trehub, 2001; Trehub & Nakata, 2002), captures and sustains attention (Nakata & Trehub, 2004), creates vital emotional bond between infant and caregiver (Shenfield, Trehub, & Nakata, 2003), and supports language development (Schön, Moreno, Boyer, Kolinsky, Besson, & Peretz, 2006) and music development (Costa-Giomi, 2013). Parents report singing more often to infants than older children and adapting their singing to the needs of the child (Custodero et al, 2003). In fact, parents sing differently to their infants and preschool children (Trehub, Trainor & Adachi). How do siblings adapt their singing when singing to infants, adults, and themselves? To answer this question, we studied the soundscapes of three families with multiple children including infants. We recorded episodes of children singing to their infant sibling, adults, and themselves (children: two girls and one boy; infants: one girl and three boys). The singing samples were gathered in the children's natural environment and produced spontaneously by the children. The singing samples varied among children in duration, content (invented songs vs known songs), lyrics (invented/accurate lyrics), and musical characteristics (e.g., loudness). Yet there was some consistency in the distinct characteristics of infant directed and adult directed singing of the three children. We discussed our findings in the context of the home soundscapes of the families taking into consideration the types of live singing to which the children were exposed.

Automated Sight-singing Assessment Tool

Helena de Souza Nunes,[*1] Rodrigo Schramm,[#2] Cláudio Rosito Jung,[#3] Lydia Wöhl Coelho[*4]

[*1] *Department of Music, UFRGS, Brazil*
[#2,#3] *Department of Informatics, UFRGS, Brazil*
[*4] *Faculty of Design, UniRitter, Brazil*

[1]helena@caef.ufrgs.br, [2]rschramm@inf.ufrgs.br, [3]crjung@inf.ufrgs.br, [4]lydiacoelho@gmail.com

ABSTRACT

Sight-singing associated with meter-mimicking gestures is often used to support the music learning process. The assessment of this practice implies the perception and feedback of each musical note, regarding parameters like pitch, onset, and offset. The traditional assessment of this practice by musician experts is very time-consuming, and can be affected by fatigue. Furthermore, since the meter-mimicking gestures may change the tempo of the music, the human perception may be drastically corrupted. This work presents a tool for automatic sight-singing assessment. The proposed system combines melodic transcription algorithms (audio analysis) with gesture tracking (video analysis) into a probabilistic model, which was built based on a statistical analysis of real human evaluations. This statistical analysis performs the mapping of similarities and differences of the audiovisual perception of musician experts to a machine learning algorithm. To accomplish this task, an external committee of musician experts conducted an annotation process along several sight-singing assessments, labeling each sung note as correct or incorrect, regarding the internal and individual decisions. The disagreements among the human evaluators in the experiments show that 10−15% of the cases have a high degree of doubt. The probabilistic approach emulates the limits of the human perception by implementing a rejection rule that discards those evaluations which are not reliable. The proposed system can perform the sight-singing assessment, note by note, in 85% of the cases, with an average accuracy of approximately 88.5%. This work provides a reliable and automated alternative for sight-singing assessment, which can be useful in the context of self-education and distance learning.

I. INTRODUCTION

Sight-singing is a usual practice used by students to improve musical reading skills through the repeated singing of musical notes from a music score. This practice is also known as solfège and when combined with meter-mimicking exercises is additionally a powerful tool for aiding the rhythmic and the sensorimotor development.

Related works have addressed this subject in parts. Regarding the body movement, Maes et al. (2013) and Schramm & Jung (2015a) developed methods for meter-mimicking assessment. Focusing in the singing analysis, algorithms have been proposed to estimate the fundamental frequency (Mauch and Dixon, 2014), pitch tracking (*Zahorian and Hu, 2008*), melodic transcription (Ryynänen, 2008) and singing assessment (Molina et al., 2013, Schramm et al. 2015b). In this context, this paper presents a combination of several techniques as an audiovisual processing approach, which performs the assessment of sight-singing exercises conducted by hand gesture movements (meter-mimicking). The main contribution of this work is the development of a statistic method for the automated assessment of this complex task, which maps the real human perception into a machine learning algorithm. This process creates a model that incorporates the indecision of expert listeners through a probabilistic (Bayesian) threshold that separates the concept of correct or incorrect.

This approach is built over a real and consistent dataset generated by a group of singers and annotated by a committee of expert listeners. The entire process creates a new method for sight-singing assessment, also regarding the possibility of *tempo* variation, which has objective and reliable metrics, being robust to situations of fatigue and subjective human interpretation.

II. METHOD

Our system uses a combination of computer vision and audio processing techniques to extract a set of features that allows the automated assessment of rhythmic and melodic performances during the study of sight-singing exercises. The audio and video information is captured during the performance of the musical exercise, and then processed in sequence by our system, which returns a visual feedback to the student. This feedback was designed to indicate, for each sung note, whether the parameters onset, duration and pitch are correct or incorrect.

A. Apparatus

The developed tool uses an RGB-D camera with a built-in array of microphones to track the hand movement and capture the audio signal. Since we had designed this tool for use in real situations of music learning in the classroom, we decided to use the MS Kinect sensor, which is an off-the-shelf RGB-D camera with low price and it is easily found in the market. The hand movement is tracked with 30 FPS. The obtained temporal 3D data point sequence is filtered to remove eventual noise. The smoothed signal is then used by a *tempo* synchronization algorithm, which also performs the musical metric pattern recognition. The audio signal is captured by a built-in array of microphones and converted to digital signal using a sample rate of 16kHz and a resolution of 16 bits. A low-pass filter is also used in a pre-processing step for removing eventual ambient noise in the audio signal. The sequence of movement as well as the audio signal are then processed by a real-time algorithm implemented in C++. This software prototype was tested in a MacBook Pro Retina with Intel Processor i7 2.4Ghz and 16 GB 1600 MHz DDR3, running Windows 7 OS. After the capture and pre-processing of audio and video signals, the data is sent to a second module of our system, which is implemented in MATLAB. This module performs the alignment between the meter-mimicking gesture and the music score (*tempo* synchronization), the melodic transcription, and the sight-singing assessment.

B. Pipeline

In contrast to previous approaches, in which the meter-mimicking evaluation is performed apart from the sight-singing assessment (Molina et al., 2013, Schramm et al. 2015b), this present approach brings up the assumption of interdependence between the kinesthesia and the vocal emission of musical notes. Thus, while a devoted part of our system extracts and evaluates the pitch/onset/duration of each sung note, another part carry out the *tempo* alignment, conducted by the meter-mimicking gestures. This complementary process allows us assessing solfège exercises even when the student changes the musical *tempo*. We structured this algorithm pipeline into the following tasks:

1) *Melodic transcription:* After capturing the audio signal from the sight-singing performance, we used the pYIN algorithm (Mauch & Dixon, 2014) to extract a sequence of fundamental frequency estimates, which are further clustered into a set of sung notes. Each detected sung note is represented as symbolic data in a tuple (onset, duration, pitch), similar to the MIDI representation.
2) *Tempo rectification and gesture recognition:* In our context, a sight-singing exercise is also defined as a list of symbolic notes (MIDI file). However, despite the onset, offset and musical *tempo* indications, there is no guarantee that the student will perform the solfège exercise in synchronization with the music score, since (s)he is allowed to control the tempo through meter-mimicking gestures. Our system implements a dynamic time warping (DTW) approach (Schramm et al., 2015a), which aligns the hand gestures with the expected musical *tempo*. Thus, the conducting movement acts as a metronome. Besides the time alignment, this task also implements a gesture classifier that is used to validate the student's meter-mimicking movement.
3) *Melodic alignment:* After the *tempo* rectification, the system makes a link from adjacent groups of sung note candidates to the respective notes that best fit the global note sequence in the music score, ensuring a mapping one-to-one. This procedure is accomplished by an optimization algorithm.
4) *Note-level assessment:* At this step, each sung note is connected with the respective most potential target note in the music score. As a result, we are able to make a direct comparison between each sung note and the linked one in the ground truth (music score). This comparison includes the evaluation of onset, duration and pitch deviations. This evaluation is implemented by a Bayesian classifier that is trained with the extracted information from real assessments done by expert listeners.

Details of this pipeline implementation are partially published in Schramm et al. (2015a, 2015b). Figure 1 illustrates the use case of the proposed system.

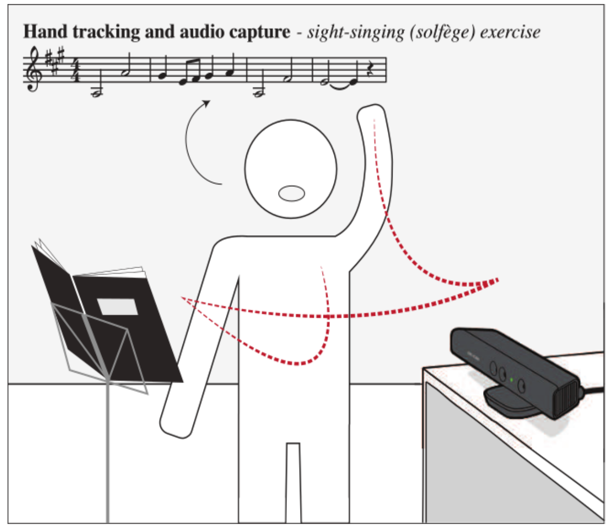


Figure 1: Automated Sight-singing assessment. Audio and video are captured by an RGB-D camera. Solfège evaluation is performed regarding the *tempo* variation through meter-mimicking conducting.

C. System Feedback

Throughout the system development, studies on user experience and usability have shown relevant to help on the design of the interactivity of the proposed tool. These studies are particularly significant to build the feedback module of our tool. Allanwood & Beare (2012, p. 14) consider user experience design (UXD) as "a collection of methods applied to the process of designing interactive experiences". Usability, in turn, is a measure of how easily a user can achieve a goal on these experiences through effectiveness, efficiency and user satisfaction (ISO 9241, 1998). UXD main focus regards websites interactivity, but also refers to the development of any digital content. Accordingly to Travis (2011), there are four principles keys to create more usable interfaces in visual design, namely: 1) *contrast*, which can be used in colors and text changes for two different actions. These changes can be trivial, like bold/italic type, color combinations, etc.; 2) *repetition*, which entails consistency by putting common icons and elements from the interface layout in standard locations, whereas users are used to follow standard templates to navigate on websites; 3) *alignment*, which requires that all elements of visual interface are lined up horizontally and vertically, transmitting an idea of easy use; and 4) *proximity*, which was first articulated by Gestalt school of psychology as one of the principles of human perception. This principle considers that, if elements are near to each other, human perception will recognized them as a group. Travis (2011) also recommend designing an interface as visible and precise as possible. In accordance with this theories, it was created the feedback layout on the system by the aid of a brief study on semiotics and color psychology.

During the automated assessment, the developed system plots the transcribed melody in a chart (Time *versus* Pitch), also showing the reference notes from the music score. Thus, it gives a fine visual pitch and timing comparison. This chart works as the feedback tool, requiring a proper layout for better interaction with the user. For Charles S. Peirce semiotics theory (Santaella, 2000), the human mind is diagrammatic and the development of knowledge occurs through sensory experiences. Therefore it was decided to build the chart as a synoptic diagram (see Figure 2), in order to stimulate the visual sense. It aims to promote a representative analysis through geometrical shapes. In this case, triangles mark the beginning and the end of a particular note (pointing up and down, respectively), and circles mark the pitch precision. The colors of these shapes indicate the accuracy of each note parameter. These colors are complementary in the RGB model and antagonic on themselves: green as a "cold" color, and red as a "warm". This contraposition is in agreement with the contrast key, suggested by Travis (2011). Since the color red is the one with longer wavelength and lower frequency on the visible chromatic spectrum, it is the color that first impact the human eye; thus, it is also the one that first reach the visual sense (Farina *et al.*, 2006, p. 92). Eva Heller (2012) considers that "warm" colors causes anxiety, heat and excitation. In this way, red color has the potential to draw attention and signalize something wrong, implying an understanding of "mistake". The green color, which is complementary and opposite to red in the RGB model, counteracts the meanings attributed to it, implying ideas of quietness and welfare on the viewer, causing the understanding of "success". When the proposed system is unable to find a confident response, it draws the respective symbol in gray color, which is a neutral tone (Farina *et al.*, 2006, p. 98). It indicates resignation or abscence of something – in this case, it means the system indecision. Figure 2 shows a print-screen of the proposed visual feedback.

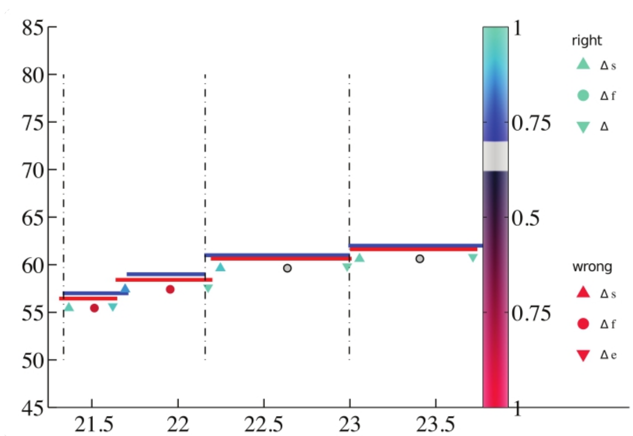


Figure 2: Visual Feedback: Triangles mark the beginning and the end of a particular note (pointing up and down, respectively), and circles mark the pitch precision. The colors of these shapes indicate the accuracy of each note parameter.

III. RESULTS

A. Dataset

The data used in this work was captured between 2013 and 2015, where sight-singing exercises were recorded by students and professors of the undergraduate music course from Federal University do Rio Grande do Sul (Brazil), providing useful input data to our experiments and evaluation. During this period, a total of 93 audio examples were recorded (based on several solfège exercises). The recording examples contains solfège exercises based on diatonic scales and all melodic intervals of the chromatic scale, associated to durations of half, quarter and eighth notes. From this initial dataset, we kept the recordings belonging to a subset of the original singers, regarding the largest intersection of examples and, at same time, the most frequent type of exercises. At the end of this screening process, our dataset contains a total of 66 recordings with examples from six distinct music scores and from fourteen distinct singers (age range from 19 to 58 years old). Later, a committee of five expert listeners, with more than ten years of experience in solfège assessment, annotated each sung note existing in the final dataset. As a result of this process, each sung note had received labels ("correct" or "incorrect") from each evaluator, regarding independently the parameters pitch, onset and offset. The final dataset contains a total of 2116 evaluated music notes (an average of 423.2 notes per expert). The ground truth is obtained for each of these sung notes by labeling each respective note parameter as correct or incorrect by the majority of votes.

It is worth noting that the evaluation was performed in two distinct contexts: a) 14% of our samples were recorded with tempo variation by meter-mimicking. Thus, its assessment is regarding the coherence between the hand position and the note onset/offset timestamp;. b) 86% of the samples were recorded in synchronization with a metronome in constant *tempo*. In this case, the assessment is performed regarding only the auditive stimulus. The disagreement among the experts were kept to model the Bayesian classifier.

B. Data Analysis

Aiming to compare the two assessment contexts, a statistical analysis was done on the voting process made by the expert's committee. We found out that, among the music notes with majority of votes for the classes "correct" and "incorrect", 10 – 15% of them received only 3 of 5 votes in agreement, regarding independently the both approaches (with and without meter-mimicking). This indicates an arguable degree of doubt among the expert listeners and also a quite consistent behavior (similar percentage), independently of the audio or audiovisual stimulus.

The final automated note-by-note assessment is performed through a Bayesian classifier that was trained using the annotated dataset, including the votes in disagreement among the experts. Figures 3 and 4 show the histograms of the deviations (distance between the measured sung note and the expected note that is defined in the music score) of the pitch, onset and offset parameters, estimated from the annotated dataset (for the correct and incorrect labels) using constant *tempo* and varying *tempo* through meter-mimicking, respectively. We have modeled the decision boundaries (Bayes Decision Rule) of the probabilistic classifiers by fitting these data into Gamma density distributions. The accuracy of the solfège assessment achieved by our system was evaluated using 10-Fold cross validation with Bayes rejection rule (Webb, 2011), discarding 15% of the samples. The final average accuracy over the three parameters is approximately 88.5%.

We also investigate the human perception tolerance through the mapping of the expert listeners' assessment, based on the Bayesian decision points for each note parameter. Figure 5 illustrates these points as well as the final trained Bayesian classifiers for the pitch, onset and offset parameters. As can be seen, the decision points (upright dashed lines) estimated using the auditive stimulus only (top) are different from the decision points estimated using the audiovisual stimulus (bottom).

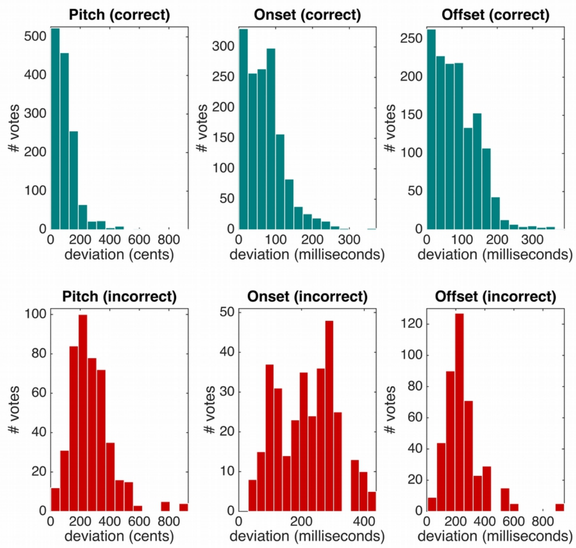


Figure 3: Estimates using the proposed system with only audio information. Histograms of the deviations between the measured sung note properties and the respective values in the music score.

We believe that the visual stimulus, together with the musical *tempo* variation, turn the human assessment more complex, making its judgment more unstable. In fact, the thresholds defined by the decision points in each note parameter are higher in the Bayesian classifier trained with audiovisual stimulus than in the classifier trained only with audio stimulus assessment. Table 1 shows the final thresholds learned by the Bayesian classifiers (edge between the correct and the incorrect classes). This means that the expert listeners were more generous in the case of the assessment with audiovisual stimulus. Surprisingly, the pitch perception was drastically affected. This can be explained by the fact that the listeners had to pay more attention on the onset and offset note parameters, which in this case have unpredictable value, since the singer might change the musical *tempo* at any time.

To validate the assumption that the expert listeners change their judgment criteria depending of the type of stimulus, we have tested the data from the two samples (audio only *versus* audiovisual) to see if they belong to the same Gamma distribution (H_0 hypothesis). A two-sample Kolmogorov-Smirnov test with 5% significance level was applied and the H_0 hypothesis was rejected (p-value $<$ 0.01 in all note parameters), indicating that the two set of data do not belong to the same distribution.

Table 1: Thresholds learned by the Bayesian Classifiers

Note Parameter	Audio Only	Audiovisual
Pitch (cents)	144	298
Onset (milliseconds)	120	142
Offset (milliseconds)	140	171

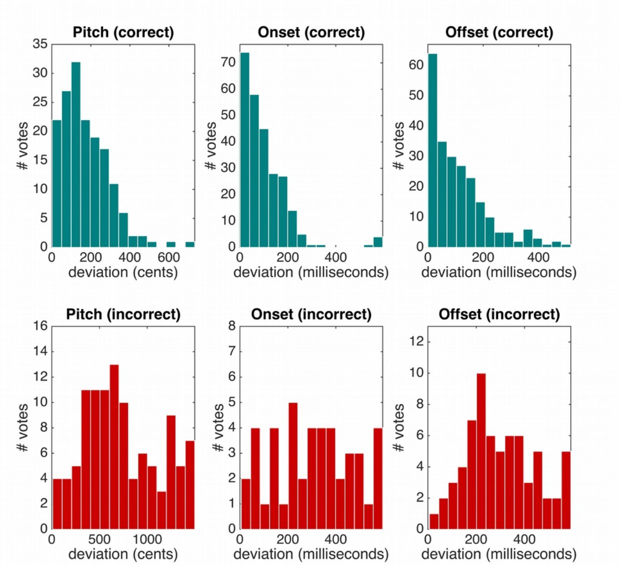


Figure 4: Estimates using the proposed system with audiovisual information (meter-mimicking). Histograms of the deviations between the measured sung note properties and the respective values in the music score.

These statistical results corroborate for the assumption that the type of stimulus have influence in the expert judgment. In the case of this application, its causes a relaxation in the human perception (threshold) between the concept of correct and incorrect. However, on the other hand, one must consider that in the first set of examples (audio only) the sung notes obey the metronome. Thus, it is possible that the relaxation of the expectation of the beat times (musical *tempo* variation) might be responsible for the distraction on the rigor of the auditive perception of the note parameters.

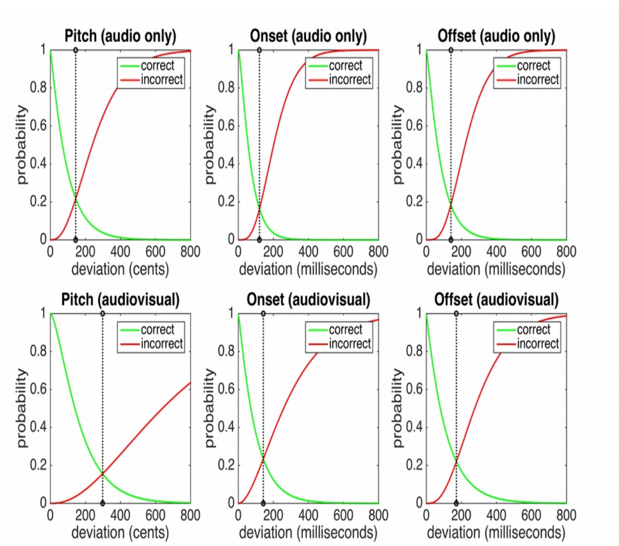


Figure 5: Bayesian Classifiers. Upright dashed lines indicate the edge between correct and incorrect classes.

IV. CONCLUSION

This work presented a tool for automated sight-singing assessment, implemented through an audiovisual approach, allowing the musician student to sing, and concomitantly, to control the music tempo with meter-mimicking gestures. The assessment is done by a set of Bayesian classifiers that create probabilistic regions of acceptance (correct class) and rejection (incorrect class), emulating the real human perception of musical parameters (pitch, onset and offset). Specifically, the Bayesian classifiers were built by fitting Gamma density distributions on the histograms obtained from measures of deviation between the detected sung note parameters and the respective note parameters specified in the music score of the sight-singing exercise. The proposed system can perform the sight-singing assessment, note by note, in 85% of the cases, with an average accuracy of approximately 88.5%. This work provides a reliable and automated alternative for sight-singing assessment, which can be useful in the context of self-education and distance learning. A visual feedback based on user experience and usability theories was proposed to aid the solfège study. It still needs more research to measure and evaluate the impact of the user interface design in the student learning process. Our experiments also have shown that the human evaluators (expert listeners) change their judgment criteria depending of the type of stimulus. Despite formal statistic tests had confirmed this behaviour, it still pending an analysis in deep to check if its cause is the difficulty of a visual analysis (audio only *versus* audiovisual) or if it is caused by the relaxation of the expectation of the musical *tempo* generated by the meter-mimicking conducting. In future work, we plan to address these questions, implementing more extensive tests based mainly on human perceptual experiments.

REFERENCES

Allanwood, G.; Beare, P. *User experience design - Creating designs users really love.* London: Bloomsbury Publishing, 2014.

Farina, M.; Perez, C.; Bastos, D. *Psicodinâmica da cores em comunicação.* São Paulo: Edgar Blücher, 2006.

Heller, E. *A Psicologia das Cores: Como as cores afetam a emoção e a razão.* 1. Ed. São Paulo: GG Editor, 2012.

ISO. *Ergonomic requirements for office work with visual display terminals (VDTs), Part 11, Guidance on usability (ISO 9241-11:1998E).* International Standards Organization, Geneva, 1998.

Maes, P., Amelynck, D., Lesaffre, M. Leman, M., and Arvind, D. K. *The "Conducting Master": An Interactive, Real-Time Gesture Monitoring System Based on Spatiotemporal Motion Templates.* International Journal Human Computer Interaction 29, 7, 2013, 471–487.

Mauch, M. and Dixon, S. 2014. *pYIN: A Fundamental Frequency Estimator Using Probabilistic Threshold Distributions.* In Proceedings of the IEEE Inter- national Conference on Acoustics, Speech, and Signal Processing (ICASSP). IEEE, Florence, 659–663.

Molina, E.; Barbancho, Gomez, E.; AM. Barbancho, and Tardon. L.J. 2013. *Fundamental frequency alignment vs. note-based melodic similarity for singing voice assessment.* In Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE Interna- tional Conference on. 744–748.

Ryynänen, M. P. and Klapuri, A. P. 2008. *Automatic transcription of melody, bass line, and chords in polyphonic music.* Computer Music Journal 32, 3 (September 2008), 72-86.

Santaella. L. A Teoria Geral dos Signos: *Como As Linguagens Significam As Coisas.* São Paulo: Pioneira Editor, 2000.

Schramm, R., Jung, C.,R., and Miranda, E. R., *Dynamic Time Warping for Music Conducting Gestures Evaluation*, Multimedia, IEEE Transactions on, vol. 17, no. 2, pp. 243-255, Feb, 2015a.

Schramm, R., H. S. de Nunes, and C. R. Jung, *Automatic Solfège Assessment*, International Society for Music Information Retrieval Conference, Málaga, 27 October, 2015b.

Stephen A. Zahorian, and Hongbing Hu, *A spectral/temporal method for robust fundamental frequency tracking*, J. Acosut. Soc. Am. 123(6), June 2008.

Travis, D. *A CRAP way to improve usability. In: Bright ideas for user experience designers.* Userfocus, 2011. Available in <http://www.userfocus.co.uk/pdf/Bright_Ideas_for_UX_Designers.pdf> Access in May, 2016.

Webb., A. R. *Statistical Pattern Recognition*. Wiley, Chichester,UK, 3 edition, 2011.

Zahorian, S.A. and Hu H. *A spectral/temporal method for robust fundamental frequency tracking.* The Journal of the Acoustical Society of America, 123, 4559-4571, 2008.

Musicians' Earplugs: Does Their Use Affect Performance or Listeners' Perceptions?

Karen S. Thomas[1], Raychl Smith[2], Sandra Teglas[1], and Donald A. Hodges[1]
[1]School of Music, Theatre and Dance, The University of North Carolina at Greensboro, Greensboro, North Carolina, USA
[2]School of Music, East Carolina University, Greenville, North Carolina, USA

The risk of noise-induced hearing loss, which results from repeated exposure to loud sounds, is a concern for those in the field of music, such as professional musicians and music educators. Many musicians may be reluctant to wear earplugs to protect their hearing, due to lack of comfort and inability to hear instrument sound properly. The purpose of this study was to determine the effect of musicians' earplugs on instrumental performance and the perception of tone quality, intonation, and dynamic contrast. Participants (N = 96) were studio faculty teachers (n = 8) and undergraduate music education students (n = 88), from a large state university school of music. Faculty teachers were recorded performing musical passages first without earplugs and then with musicians' earplugs. Objective data were collected to compare recorded frequencies with standard frequencies of single pitches recorded with and without earplugs. A perception test was created using individual audio tracks with pairs of the same performer and performance material, with and without musicians' earplugs. Participants listened to each pair of phrases to determine whether the second phrase differed from the first in regards to tone quality, intonation, and dynamic contrast. Acoustical analyses of recordings made by faculty with and without earplugs indicate that pitch accuracy did not consistently favor either condition. Preliminary results from the perception test indicate that although both faculty and student listeners perceived some differences, the most frequent perception was that the audio pair was equal, and there was no clear advantage between performing with and without earplugs in terms of tone quality, intonation, and dynamic contrast. The implications of these findings suggest that musicians could feel confident that wearing musicians' earplugs while performing may not adversely affect pitch accuracy or listeners' perceptions of their timbre and dynamic control.

Student persistence in group-based Suzuki music training and parent ratings of children's empathy

Kathleen M. Einarson[1], Karin S. Hendricks[2], Nancy Mitchell[3], Elizabeth M. Guerriero[4], Patricia D'Ercole[5]
[1] Department of Psychology, Neuroscience & Behaviour, McMaster University, Hamilton, Ontario, Canada
[2] College of Fine Arts, Boston University, Boston, Massachusetts, USA
[3] Department of Music, Durham School Board, Pickering, Ontario, Canada
[4] Department of Music, Montgomery County Community College, Blue Bell, Pennsylvania, USA
[5] Aber Suzuki Center, University of Wisconsin – Stevens Point, Stevens Point, Wisconsin, USA

Shinichi Suzuki emphasized fostering kindness and sensitivity to others as part of music instruction, and group music making impacts children's empathy and social skills. We investigated whether six months of Suzuki group music lessons increased children's empathy, as assessed by the Griffith Empathy Measure (GEM). Parents of students in a Suzuki instrumental music program ($N = 48$) completed the GEM at the start of their lesson year, and again six months later, and were also surveyed about their values related to music and family sociodemographic data at Time 1. A 2x3 (empathy score x group persistence) ANOVA revealed no significant change in empathy from Time 1 to Time 2. However, we noted differing trends, consistent with our predictions, among those who did or did not continue group lessons: Mean empathy scores for children who stayed in lessons increased from $M = 2.44$ to $M = 2.67$, while scores for those who did not persist (or did not participate at all) decreased from $M = 1.71$ to $M = 1.58$ over the same period. The analysis revealed a significant between-subjects effect for group participation ($p = .004$), and post-hoc analyses showed a significant difference in Time 1 empathy scores between children who subsequently persisted in group class and those who never enrolled ($p = .003$). These pre-existing differences in our sample are consistent with Corrigall and Schellenberg's (2015) observation that child personality predicts music training duration. Children's empathy score at Time 1 was also significantly associated with parent's rating that music was important ($\rho = .54$, $p = .002$), and that Suzuki instruction in particular was important ($\rho = .54$, $p = .003$), consistent with Corrigall and Schellenberg's finding that music training is predicted by parent personality (2015). We conclude that individual differences play an important role in music education, and continue to explore group music and empathy in an ongoing study of Suzuki programs across North America.

Cognitive control in musicians: does the task matter?

Marie-Eve Joret[1], Kobe Desender[2]
[1]Department of Linguistics, Vrije Universiteit Brussel, Brussels, Belgium
[2]Department of Psychology, Vrije Universiteit Brussel, Brussels, Belgium

Background
Music training has been associated with enhanced cognitive control, which refers to the cognitive processes enabling self-regulation of behavior. However, the precise association between music training and cognitive control remains unclear. Possibly, this stems from studies using different types of tasks that measure different aspects of cognitive capacities (linguistic, visuospatial, etc.) or that rely on different response modes.

Aims
We investigated whether the enhanced performance by musicians on cognitive control tasks depends on the type of task.

Method
83 participants were tested (9-11 years old), with 39 monolingual musically trained children and 44 monolingual non-musically trained children. Information on confounders, such as socio-economic status, verbal and non-verbal IQ and personality were measured to ensure group equality. To assess cognitive control, both groups underwent the Simon and the Stroop task, measuring reaction times (RTs) and error rates on congruent and incongruent trials.

Results
There were no differences between groups on the background variables. Concerning the RTs on the Simon task, a significant interaction of group by congruency emerged (one-tailed), such that musicians had a smaller interference effect than controls. For the error rates, there were no differences between groups. Concerning the Stroop task, no differences between the musicians and the controls were found on the RTs and the error rates.

Conclusion
Our results suggest that the cognitive control benefit for musicians could depend on the cognitive control task. A potential explanation for these findings is that musicians are better at processing spatial information, which is an important factor in the Simon task, but not in the Stroop task.

Specific Mathematical and Spatial Abilities Correlate with Music Theory Abilities

Nancy Rogers,[*1] Jane Piper Clendinning,[*2] Sara Hart,[#3] and Colleen Ganley[#4]

[*]*College of Music, Florida State University, USA*

[#]*Department of Psychology, Florida State University, USA*

[1]nancy.rogers@fsu.edu, [2]jclendinning@fsu.edu, [3]shart@fcrr.org, [4]ganley@psy.fsu.edu

ABSTRACT

There are widespread beliefs in links between mathematics and music. Such connections have been asserted at least since the time of the Ancient Greeks, and contemporary music theory teachers anecdotally report that students who experience extraordinary difficulty grasping material in their music theory classes typically also struggle in their math classes (and sometimes contend with a diagnosed mathematical learning disability). Most previous studies of this phenomenon have addressed broad mathematical abilities and broad musical abilities (both of which are difficult to distinguish from general intelligence). Our aim is to demonstrate more specific connections between the skills addressed in introductory music theory courses and particular mathematical and spatial abilities.

In our study, 99 trained musicians in the initial week of a first-semester college-level music theory course completed a battery of tests evaluating their sense of number, mental rotation ability, geometric and musical pattern recognition, and levels of confidence and anxiety regarding both mathematics and music theory. Participants' music theory grades were obtained over the academic year. We observed numerous significant positive correlations between mathematical/spatial skills and music theory performance; most notable among these were musical pattern recognition, geometric pattern recognition, and mental rotation. There were also significant positive correlations between confidence in both music theory and mathematics and performance in music theory classes, as well as significant negative correlations between anxiety in both music theory and mathematics and performance in music theory classes. Although much work remains, we believe that our results greatly clarify some components of the oft-cited connection between mathematics and music

I. INTRODUCTION

Experienced music theory teachers become concerned when struggling students express antipathy towards mathematics, describe themselves as incompetent at mathematics, or display significant anxiety about mathematics. It is widely accepted among music theory instructors that mathematical ability (or, more precisely, mathematical inability) strongly predicts basic written music theory skills. Such observations are by no means recent; there are long-standing beliefs that musical abilities correlate with mathematical abilities. The Ancient Greeks, for instance, considered music to be a form of mathematics. Sixth-century Romans grouped music with arithmetic, geometry, and astronomy and called this set of four subjects the quadrivium. More recently, Stravinsky said that music is "something like mathematical thinking and mathematical relationships" (Craft, 1959). As indicated earlier, music theory students routinely claim that music theory is like math. By this point, a music-math relationship probably qualifies as conventional wisdom.

Numerous scholars have provided supporting evidence for these assumed links between music and math. Some have framed the issue within the broader context of general intelligence, discussing music's impact on a variety of other abilities, including mathematics (e.g., Schellenberg 2001, 2004, 2005, 2006, 2011). Those that focus more narrowly on music and mathematics (e.g., Schmithorst & Holland, 2004; Haimson, Swain, & Winner, 2011; Cox & Stephens, 2004; Helmrich, 2010; Bahr & Christensen, 2000; Brochard, Dufour, & Després, 2004; Cheek & Smith, 1999) typically report positive correlations between mathematical and musical ability and/or achievement. The vast majority of music/math research compares musicians with non-musicians, and the musical focus tends to be on performance. The notable exceptions are Bahna-James (1991), Harrison (1990a, 1990b, 1996), and Rogers & Clendinning (2015), all of whom demonstrated correlations between students' mathematical abilities and their performance in introductory music theory courses. While math/music studies generally report that superior musical ability or high musical achievement correlates with superior mathematical ability or high mathematical achievement, Rogers & Clendinning also provided evidence that math disabilities correlate with unusual difficulties in music theory.

Although there is considerable support for the hypothesis that there is some link between mathematical and musical abilities, the specific nature of this link is unclear. It appears unlikely that the association is causal: combined experimental work indicates that random assignment to music training has only a small positive effect on math achievement, and some find this evidence unconvincing due to a lack of available studies (Vaughn, 2000; see also Mehr et al., 2013). It seems more probable that shared etiological factors underlie success in both mathematics and music (e.g., Jonides, 2008). Two main sources have been highlighted: cognitive processing factors and affective factors. Related to cognitive processing factors, aspects such as an understanding of mathematical relations (e.g., ratios, which are important for rhythmic proportions) and spatial reasoning (e.g., identifying a key signature quickly or picturing a triad in close position) may contribute to success in both math and music (see Schellenberg, 2011, for a review). There has been very little examination of affective aspects underlying math achievement and music success beyond general "emotional processing" (Schellenberg & Mankarious, 2012). However, the importance of math anxiety on math achievement has recently been highlighted (e.g., Beilock & Maloney, 2015); it is reasonable to wonder whether aspects such as state anxiety and confidence might also associate mathematics and music. Our study aims to clarify particular capabilities that may contribute to success in both mathematics and music.

II. METHOD

Rather than relying solely on broad measures such as standardized exams to represent math ability, we administered a battery of tests addressing basic numeracy, spatial skills, and pattern recognition, as well as surveys addressing anxiety and confidence. Just as importantly, we didn't assume that musical ability is well represented by years of private study, hours of practice time, or public concerts. We focused instead on performance in music theory classes and on music theory exams, because — as an academic study of music — music theory grades are particularly likely to reflect the underlying cognitive processing and affective factors that may link achievement in mathematics and music.

A. Participants

This study included 99 participants (60 female, 38 male; one declined to answer), all of whom were recruited from beginning music theory courses at Florida State University. Most (86) were enrolled in Music Theory I, the first course in the required four-semester music theory sequence required for music majors; the remaining 13 were enrolled in Fundamentals for Music Majors, a remedial course that precedes Music Theory I. Their mean age was 18.4 (*SE* = .12). Participants granted permission to obtain their music theory grades, entrance exam scores, and standardized test scores. (Data from several additional participants who withheld their permission have not been included in this report.) They were awarded extra credit for their participation.

B. Procedure and Apparatus

The skills tests and surveys were administered online through Qualtrics. Participants completed all components from their own computers during the first week of the semester (before the courses in which they were enrolled addressed any skills relevant to the tasks involved). Following the skills tests and surveys, participants answered a questionnaire addressing their musical background, training, and demographic information. Although many individual components were timed, students were allowed to pause between components. Not including any such pauses, the total time required was approximately 45 minutes.

C. Measures

1) Mental Rotation Test. This test (Vandenberg & Kuse, 1978) consisted of 24 items. Each item included a picture of a three-dimensional object on the left called the target object. On the right, there were pictures of four three-dimensional objects. Two of these pictures depicted the same object as the target, only rotated and thus presented from a different perspective. The other two pictures were different: they were either a mirror image of the target object or had different features. Participants were asked to identify which two of the four objects were the same as the target. A sample is provided in Figure 1.

2) Geometric Invariants Test. This test (Dehaene, Izard, Pica, & Spelke, 2006) consisted of 14 items drawn from the original 45-item scale. Each item included an array of six similar images, five of which instantiated a particular concept (e.g., symmetry or equidistance) while the remaining one violated it. Participants were asked to identify the image that didn't match the other five. Samples are provided in Figure 2.

3) Approximate Number System Test. Participants completed the 120-item Panamath Test (www.panamath.org), a freely available web-based test of the Dots Task, a measure of the approximate number system (Halberda, Mazzocco, & Feigenson, 2008). The Dots Task is a method for measuring the ability of an individual to understand and manipulate numerical quantities non-symbolically. The process underlying this ability of non-symbolic number sense is called the Approximate Number System, an intuitive recognition of number. Participants were shown collections of intermixed blue and yellow dots, with five to twenty dots per color, for 600 ms and determined whether there was a larger number of yellow dots or a larger number of blue dots. They were instructed to ignore the size of the dots, focusing solely on the quantity. A representative image is provided in Figure 3. The accuracy and response time for each trial was recorded, and this was converted into a participant's Weber fraction (w-score), which represents the smallest ratio that can accurately be discriminated by a given individual.

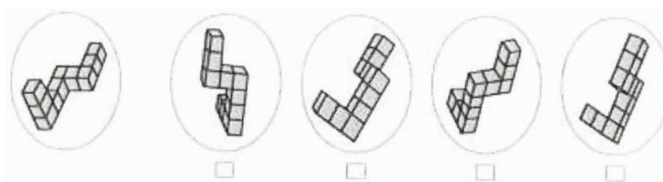


Figure 1. A sample item from the Mental Rotation Test.

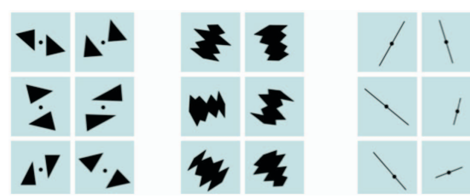


Figure 2. Three sample items from the Geometric Invariants Test.

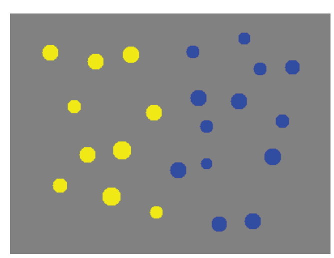


Figure 3. A sample image from the Approximate Number System Test.

4) Berlin Numeracy Test. This test (Cokely, Galesic, Schulz, Shazal, & Garcia-Retamero, 2012) consisted of four items drawn from the original collection. The Berlin Numeracy Test quickly assesses statistical numeracy and risk literacy. It has been shown to be a strong predictor of an individual's comprehension of everyday risks and statistical probabilities. Participants answered questions such as the one below.

> Imagine we are throwing a five-sided die 50 times. On average, out of these 50 throws how many times would this five-sided die show an odd number (1, 3 or 5)?

5) Standard Music Notation Recognition Test. This researcher-developed test consisted of 28 items falling into three distinct categories: individual symbols, key signatures, and rhythmic notation. For individual symbols, participants saw images of common musical symbols for 500 ms. Some images were correct, and others had been distorted (e.g., written backwards). Participants were asked to determine whether or not the image represented correct standard musical notation. Samples are provided in Figure 4.

For key signatures, participants saw images for 500 ms. Some images were correct, and others had been distorted (e.g., accidentals were written in the wrong order or in the wrong register). Participants were asked to determine whether or not the image represented correct standard musical notation. Samples are provided in Figure 5.

For rhythmic notation, participants saw images of a one-measure rhythm in either 3/4 or 6/8 for 500 ms. Some images beamed the eighth-notes correctly, while others either beamed inappropriate combinations of eighth-notes. Participants were asked to determine whether or not the image represented correct standard musical notation. Samples are provided in Figure 6.

6) Triad Recognition Test. This researcher-developed test consisted of 10 items. Because some participants had no prior instruction in music theory, they were first provided with an explanation of a triad:

> A triad contains three notes that can be arranged on a staff to fall on consecutive lines or consecutive spaces. This might involve changing octaves.

Clarifying illustrations showed a triad built on consecutive spaces, a triad built on consecutive lines, and two demonstrations in which notes were moved an octave in order to reveal the triadic structure. Participants then saw images of three pitches (without accidentals) on a single clef for 500 ms. Some of the images formed triads and others did not. Participants were asked to determine whether or not the images contained triads. Samples are provided in Figure 7.

7) Musical Pattern Continuation Test. This researcher-developed test consisted of six items. In this untimed test, participants were shown a melodic pitch pattern containing 7-10 notes. They were asked to determine which one of four possible continuations maintained the established pattern. A sample is provided in Figure 8.

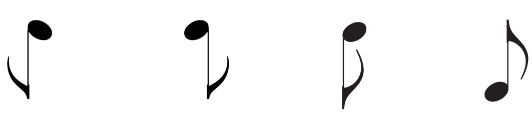


Figure 4. Four sample items from the individual symbol portion of the Standard Music Notation Recognition Test.

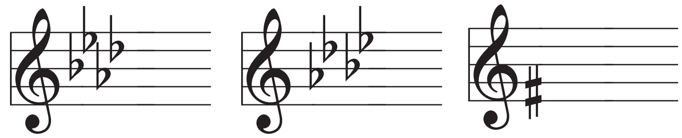



Figure 5. Three sample items from the key signature portion of the Standard Music Notation Recognition Test.

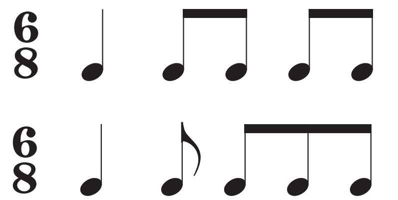


Figure 6. Two sample items from the rhythmic notation portion of the Standard Music Notation Recognition Test.

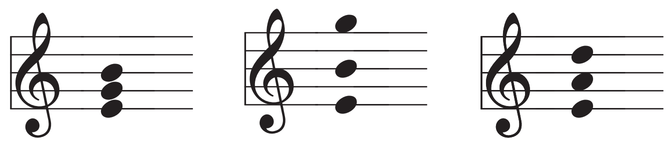


Figure 7. Three sample items from the Triad Recognition Test.

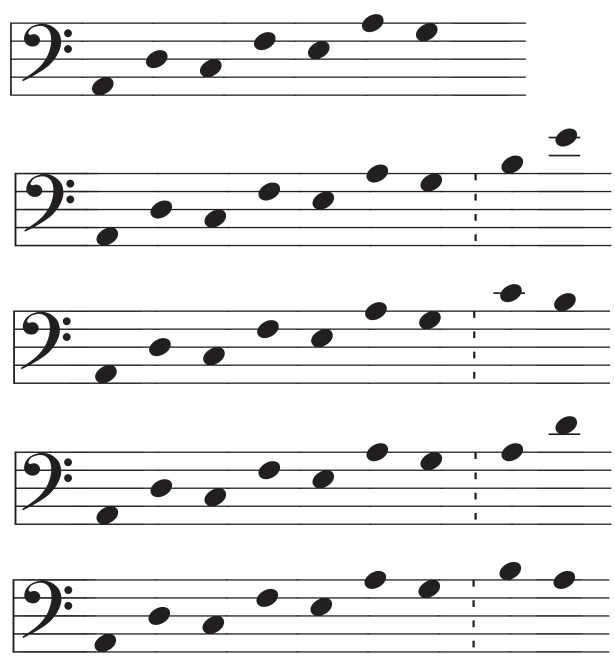


Figure 8. A sample item from the Musical Pattern Continuation Test.

8) Math Confidence Assessment. This assessment (adapted from the Confidence Subscale of the Fennema Sherman Math Attitudes Scale, Fennema & Sherman, 1976) consisted of seven statements. Participants were asked to indicate how much they agreed or disagreed with each statement on a seven-point scale (1 = "strongly disagree" and 7 = "strongly agree"). Two sample statements appear below.

I am sure I could do advanced work in math.
I am not good at math.

9) Math Anxiety Assessment. This assessment (Math Anxiety Rating Scale-Revised, Hopko, 2003) consisted of twelve situation descriptions. Participants were asked to indicate how much anxiety they would feel in each situation on a five-point scale (1 = "low anxiety" and 5 = "high anxiety"). Two samples appear below.

Thinking about an upcoming math test one day before
Watching a teacher work an algebraic on the blackboard

10) Music Theory Confidence Assessment. This assessment (adapted from the Confidence Subscale of the Fennema Sherman Math Attitudes Scale, Fennema & Sherman, 1976) consisted of seven statements. Participants were asked to indicate how much they agreed or disagreed with each statement on a seven-point scale (1 = "strongly disagree" and 7 = "strongly agree"). Two sample statements appear below.

I am sure I could do advanced work in music theory.
I am not good at music theory.

11) Music Theory Anxiety Assessment. This assessment (adapted from Math Anxiety Rating Scale-Revised, Hopko, 2003) consisted of twelve situation descriptions. Participants were asked to indicate how much anxiety they would feel in each situation on a five-point scale (1 = "low anxiety" and 5 = "high anxiety"). Two samples appear below.

Thinking about an upcoming music theory test one day before
Watching a teacher write counterpoint on the blackboard

III. RESULTS AND DISCUSSION

Very few measures correlated significantly with overall performance in the Music Fundamentals class, a remedial course for music majors who enter Florida State University with very weak backgrounds. In contrast, nearly all of our measures (with the sole exception of the Approximate Number System Test) exhibited the expected significant correlation with overall performance in Music Theory 1. Although this might suggest that our measures are better predictors for more experienced students, a more likely explanation is that an insufficient number of Fundamentals students (13) volunteered to participate in the study. Most of these Fundamentals students are, in fact, part of the Music Theory 1 data pool because they advanced to Music Theory 1 in the second semester (while most of the other 86 participants advanced to Music Theory 2).

Table 1 presents a summary of the observed correlations between individual measures and participant performance both in specific courses and more generally in all core music theory courses so far (i.e., Music Theory 1 and 2, because none of these participants have yet taken Music Theory 3 and 4). For purposes of comparison, Table 1 also includes results from the Fundamentals Placement Exam, a traditional achievement test that assesses knowledge of scales, key signatures, intervals, and triads. (The Fundamentals Placement Exam determines whether students begin with Fundamentals for Music Majors or with Music Theory 1.)

A. Individual Measures

1) Mental Rotation Test. Participant scores ranged from 1-21, out of a maximum possible score of 24, with a mean of 10.5 (*SE* = .6). The Mental Rotation Test correlated significantly with Music Theory 1, with overall performance in music theory, and with the traditional Fundamentals Placement Exam.

2) Geometric Invariants Test. Participant scores ranged from 3-14, the maximum possible score, with a mean of 11.6 (*SE* = .3). The Geometric Invariants Test correlated significantly with Music Theory 1, Music Theory 2, overall performance in music theory, and with the traditional Fundamentals Placement Exam.

3) Approximate Number System Test. Two scores were eliminated due to technical errors. The remaining scores ranged from 0.12-0.67, with a mean of 0.242 (*SE* = .01). The Approximate Number System Test did not correlate significantly with any measure of achievement in music theory. Although sense of number is reported to be an important factor in mathematical ability, it does not appear critical to music theory ability.

4) Berlin Numeracy Test. Participant scores ranged from 1-4, the maximum possible score, with a mean of 2.0 (*SE* = .1). The Berlin Numeracy Test correlated significantly with Music Theory 1, Music Theory 2, overall performance in music theory, and with the traditional Fundamentals Placement Exam. Although these correlations were less strong than those observed for the Geometric Invariants Test, the Mental Rotation Test, the Triad Recognition Test, and the Musical Pattern Continuation Test, it is worth noting that this measure included only four items.

5) Standard Music Notation Recognition Test. Participant scores ranged from 14-26, the maximum possible score, with a mean of 21.5 (*SE* = .3). Scores on individual symbols ranged from 1-4, the maximum possible score, with a mean of 3.6 (*SE* = .1). Scores on key signatures ranged from 4-10, the maximum possible score, with a mean of 8.6 (*SE* = .1). Scores on rhythmic notation ranged from 4-12, the maximum possible score, with a mean of 9.3 (*SE* = .2). The Standard Music Notation Recognition Test correlated strongly with Music Theory 1, Music Theory 2, overall performance in music theory, and with the traditional Fundamentals Placement Exam. In fact, this measure alone predicted student success in all three music theory courses better than the traditional Fundamentals Placement Exam (results for the Music Fundamentals course fell just short of significance, $p = .08$).

6) Triad Recognition Test. Participant scores ranged from 1-9, out of a maximum possible score of 10, with a mean of 5.8 (*SE* = .2). The Triad Recognition Test correlated with performance in all music theory courses, including a fairly strong significant correlation with Fundamentals. It also correlated with the traditional Fundamentals Placement Exam (which itself did not correlate significantly with performance in the Fundamentals course).

7) Musical Pattern Continuation Test. Participant scores ranged from 0, the minimum possible score, to 6, the maximum possible score, with a mean of 3.8 (SE = .2). The Musical Pattern Continuation Test was our single most successful measure, correlating with performance in all music theory courses, including a fairly strong significant correlation with Fundamentals. It also correlated with the traditional Fundamentals Placement Exam (which, as mentioned above, did not correlate significantly with performance in the Fundamentals course). The results associated with this measure are especially remarkable because it only included six items.

8) Math Confidence Assessment. Participant scores ranged from 1, the minimum possible score, to 6.7, out of a maximum possible score of 7, with a mean of 4.4 (SE = .2). The Math Confidence Assessment correlated positively with Music Theory 1, Music Theory 2, overall performance in music theory, and with the traditional Fundamentals Placement Exam.

9) Math Anxiety Assessment. Participant scores ranged from 1, the minimum possible score, to 5, the maximum possible score, with a mean of 2.4 (SE = .1). The Math Anxiety Assessment correlated negatively (that is, lower anxiety corresponded with better outcomes) with Music Theory 1, Music Theory 2, and overall performance in music theory. These negative correlations were considerably less strong than the positive correlations observed for the Math Confidence Assessment.

10) Music Theory Confidence Assessment. Participant scores ranged from 1, the minimum possible score, to 7, the maximum possible score, with a mean of 4.4 (SE = .1). The Music Theory Confidence Assessment correlated positively with Music Theory 1, Music Theory 2, overall performance in music theory, and with the traditional Fundamentals Placement Exam. These correlations were considerably stronger than those observed for the corresponding Math Confidence Assessment. One possible explanation is that most incoming music majors at Florida State University have previously studied music theory and therefore have realistic expectations for their upcoming coursework in music theory.

Table 1. Correlations between individual tests / assessments and participants averages in music theory courses. The Overall Core Theory Average combines Music Theory 1 and 2. For purposes of comparison, the traditional Fundamentals Placement Exam and a combined measure that includes all notation-based measures (the Music Notation Recognition Test, Triad Recognition Test, and Musical Pattern Continuation Test) are also included.

	Fundamentals Placement Exam	Fundamentals Course Average	Music Theory 1 Average	Music Theory 2 Average	Overall Core Theory Average
Mental Rotation	r = .225 p < .04	*n.s.*	r = .284 p < .01	*n.s.*	r = .247 p < .02
Geometric Invariants	r = .328 p < .01	*n.s.*	r = .296 p < .01	r = .242 p < .05	r = .262 p < .01
Approximate Number System	*n.s.*	*n.s.*	*n.s.*	*n.s.*	*n.s.*
Berlin Numeracy	r = .224 p < .04	*n.s.*	r = .229 p < .03	*n.s.*	r = .206 p < .05
Music Notation Recognition	r = .565 p < .0001	*n.s.*	r = .614 p < .0001	r = .464 p < .0001	r = .551 p < .0001
Triad Recognition	r = .496 p < .0001	r = .598 p < .03	r = .448 p < .0001	r = .254 p < .04	r = .360 p < .001
Musical Pattern Continuation	r = .588 p < .0001	r = .599 p < .03	r = .530 p < .0001	r = .302 p < .02	r = .470 p < .0001
Math Confidence	r = .209 p < .05	*n.s.*	r = .361 p < .001	r = .354 p < .01	r = .377 p < .0001
Math Anxiety	*n.s.*	*n.s.*	r = -.260 p < .02	r = -.257 p < .03	r = -.265 p < .01
Music Theory Confidence	r = .613 p < .0001	*n.s.*	r = .458 p < .0001	r = .395 p < .001	r = .455 p < .0001
Music Theory Anxiety	r = -.361 p < .001	*n.s.*	r = -.343 p < .001	r = -.287 p < .02	r = -.344 p < .001
Fundamentals Placement Exam	—	*n.s.*	r = .549 p < .0001	r = .460 p = .0001	r = .452 p < .0001
Combined Music Notation Measures	r = .695 p < .0001	r = .715 p < .01	r = .714 p < .0001	r = .485 p < .0001	r = .624 p < .0001

11) Music Theory Anxiety Assessment. Participant scores ranged from 1, the minimum possible score, to 4.8, out of a maximum possible score of 5, with a mean of 2.3 (SE = .1). The Music Theory Anxiety Assessment correlated negatively (that is, lower anxiety corresponded with better outcomes) with Music Theory 1, Music Theory 2, and overall performance in music theory. As observed with the corresponding Math Confidence/Anxiety Assessments, the correlations with performance in music theory courses were stronger for confidence than for anxiety. The one exception was an especially strong negative correlation with the traditional Fundamentals Placement Exam. As noted above, most incoming music majors at Florida State University have previously studied music theory. For most students who experience unusual difficulties with music theory, signs of trouble arise quite early, and so it is not surprising that these students already have high levels of music theory anxiety, whereas most students find introductory music theory courses quite easy.

B. General Discussion

Consistent with previous research (e.g., Rogers & Clendinning, 2015), math subscores of both the ACT and the SAT correlated fairly strongly with music theory grades for these participants (r = .603 and .602, respectively; $p < .0001$). Indeed, they predicted success better than did the traditional Fundamentals Placement Exam. English subscores for both the ACT and the SAT also correlated significantly with music theory grades (r = .400 and .386, respectively; $p < .01$). However, no established test outperformed the combination of our three measures that relied on music notation: the Music Notation Recognition Test, Triad Recognition Test, and Musical Pattern Continuation Test.

Interestingly, correlations between standardized math subscores and our specifically math-related measures (the Berlin Numeracy Test and the Approximate Number System Test) were not as strong as the correlations with music theory performance, nor were they always significant. In contrast, correlations between standardized math subscores and measures focusing on pattern recognition (the Geometric Invariance Test and the Musical Pattern Continuation Test) were fairly strong and invariably significant. Pattern recognition tests did not correlate with ACT English subscores, and only weakly with SAT Writing subscores.

Although we expected that the Triad Recognition Test might correlate strongly with the Mental Rotation Test, this was not the case. However, the Geometric Invariants Test correlated significantly both with the Triad Recognition Test and with the Mental Rotation Test. Not surprisingly, the Musical Pattern Continuation Test correlated significantly with the Geometric Invariants Test.

IV. CONCLUSION

Our data supports the theory that shared etiological factors underlie success in both mathematics and music. In particular, we see evidence that pattern recognition abilities and affective factors may play important roles in mathematical and musical learning.

REFERENCES

Bahna-James, T. (1991). The relationship between mathematics and music: Secondary school student perspectives. *The Journal of Negro Education 60:3*, 477–485.

Bahr, N. & Christensen, C.A. (2000). Inter-domain transfer between mathematical skill and musicianship. *Journal of Structural Learning and Intelligence Systems, 14*, 187–197.

Beilock, S.L., & Maloney, E.A. (2015). Math anxiety: A factor in math achievement not to be ignored. *Policy Insights from the Behavioral and Brain Sciences, 2:1*, 4–12.

Brochard, R., Dufour, A., & Després, O. (2004). Effect of musical expertise on visuospatial abilities: Evidence from reaction times and mental imagery. *Brain and Cognition, 54*, 103–109.

Cheek, J.M. & Smith, L.R. (1999). Music training and mathematics achievement. *Adolescence, 34*, 759–761.

Cokely, E.T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin Numeracy Test. *Judgment and Decision Making, 7*, 25-47.

Cox, H.A. & Stephens, L.J. (2004). The effect of music participation on mathematical achievement and overall academic achievement of high school students. *International Journal of Mathematical Education in Science and Technology, 37:7*, 757-763.

Craft, R. & Stravinsky, I. (1959). *Conversations with Igor Stravinsky*. Garden City, NY: Doubleday.

Dehaene, S., Izard, V., Pica, P., & Spelke, E. (2006). Core knowledge of geometry in an Amazonian indigene group. *Science*, *311:5759*, 381–384.

Fennema, E., & Sherman, J. (1976). Mathematics attitudes scales: Instruments designed to measure attitudes toward the learning of mathematics by females and males. *Journal of Research in Mathematics Education, 7*, 324–326.

Haimson, J., Swain, D., & Winner, E. (2011). Do mathematicians have above average musical skill? *Music Perception, 29:2*, 203–213.

Halberda, J., Mazzocco, M.M.M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature, 455*, 665–668.

Harrison, C.S. (1990a). Predicting music theory grades: The relative efficiency of academic ability, music experience, and musical aptitude. *Journal of Research in Music Education, 38:2*, 124–137.

Harrison, C.S. (1996). Relationships between grades in music theory for nonmusic majors and selected background variables. *Journal of Research in Music Education, 44:4*, 341–352.

Harrison, C.S. (1990b). Relationships between grades in the components of freshman music theory and selected background variables. *Journal of Research in Music Education, 38:3*, 175–186.

Hart, S.A., Logan, J.A.R., Thompson, L.A., Kovas, Y., McLoughlin, G. & Petrill, S. A. (in press). A latent profile analysis of math achievement, numerosity, and math anxiety in twins. *Journal of Educational Psychology*.

Helmrich. B.H. (2010). Window of opportunity? Adolescence, music, and algebra. *Journal of Adolescent Research 25:4*, 557–577.

Hopko, D.R. (2003). Confirmatory factor analysis of the math anxiety rating scale–revised. *Educational and Psychological Measurement, 63:2*, 336–351.

Jonides, J. (2008). Musical skills and cognition. *Learning, Arts, and the Brain: The Dana Consortium Report on Arts and Cognition*, 11–16.

Kenny, D.T., Davis, P., & Oates, J. (2004). Music performance anxiety and occupational stress amongst opera chorus artists and their relationship with state and trait anxiety and perfectionism. *Journal of Anxiety Disorders, 18:6*, 757–777.

Mehr S.A., Schachner, A., Katz, R.C., & Spelke, E.S. (2013). Two randomized trials provide no consistent evidence for nonmusical

cognitive benefits of brief preschool music enrichment. *PLOS ONE, 8*, e82007.
Muthén, L.K., & Muthén, B.O. (2007). Mplus User's Guide. *Journal of the American Geriatrics Society, 2006*, 676.
Rogers, N.M., & Clendinning, J.P. (2015). Music Theory Ability Correlates with Mathematical Ability. Meeting of the Society for Music Perception and Cognition, Nashville, TN.
Schellenberg, E.G. (2011). Examining the association between music lessons and intelligence. *British Journal of Psychology*, *102*, 283–302.
Schellenberg, E.G. (2006). Long-term positive associations between music lessons and IQ. *Journal of Educational Psychology*, *98:2*, 457–468.
Schellenberg, E.G. (2005). Music and cognitive abilities. *Current Directions in Psychological Science, 14*, 322–325.
Schellenberg, E.G. (2004). Music lessons enhance IQ. *Psychological Science, 15*, 511–514.
Schellenberg, E.G. (2001). Music and nonmusical abilities. *Annals of the New York Academy of Sciences, 930*, 355–371.
Schellenberg, E.G., & Mankarious, M. (2012). Music training and emotion comprehension in childhood. *Emotion, 12:5*, 887.
Schmithorst, V.J. & Holland S.K. (2004). The effect of musical training on the neural correlates of math processing: A functional magnetic resonance imaging study in humans. *Neuroscience Letters, 354*, 193–196.
Vandenberg, S.G., & Kuse, A.R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 47*, 599–604.
Vaughn, K. (2000). Music and mathematics: Modest support for the oft-claimed relationship. *Journal of Aesthetic Education*, *34:3/4*, 149–166.

The Effect of Movement Instruction in Music Education on Cognitive, Linguistic, Musical and Social skills

Emese Maróti[1,2], Edina Barabás[3], Gabriella Deszpot[3], Tamara Farnadi[4], László Norbert Nemes[3], Borbála Szirányi[3] and Ferenc Honbolygó[1]
[1]Brain Imaging Centre, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest, Hungary
[2]Department of Cognitive Science, Budapest University of Technology and Economics, Budapest, Hungary
[3]Kodály Institute, Liszt Ferenc Academy of Music, Budapest, Hungary
[4]János Richter Conservatory of Music, Győr, Hungary

Several studies have shown the benefits of music training on cognitive, linguistic and neural development. It has been also found that incorporating movement instruction into music education can improve musical competence. However, it is not known whether movement instruction in music education can also relate to other, non-musical skills. Here we address this question and report the results of the first year of a longitudinal study, in which we examine the effect of movement instruction in childhood music education on musical, cognitive, linguistic and social development. We compared 6-7 year old children receiving the traditional Kodály music training, which includes solfeggio and singing, with two groups of children receiving movement based music training: one involves rule based sensorimotor synchronization, and the other involves free improvisation (Kokas method). In order to measure the effect of these methods, we created a comprehensive test battery for sensorimotor synchronization, musical, cognitive, linguistic and social skills. In the beginning of the study children were compared on all tests to examine any possible pre-existing differences between the groups that can affect the results of future comparisons. Our results revealed that the groups did not differ initially in any musical and non-musical skills. After 8 months, children underwent the same tests, and we found that movement based methods improved sensorimotor synchronization, rhythmic ability, linguistic skills and executive functions. Furthermore, free improvisation increased pitch discrimination. Finally, we found significant correlations between sensorimotor synchronization and both executive functions and linguistic skills. These initial results inspire us to continue investigating whether the effect of movement instruction can manifest not only on a behavioral, but on a neural level, specifically how it affects neural entrainment to a musical rhythm and neural encoding of speech.

Singing for the Self: exploring the self-directed singing of three and four year-old children at home

Bronya Dean[1]
[1]Graduate School of Education, University of Exeter, Exeter, United Kingdom

The everyday musical lives of young children are a subject of growing interest among music education researchers and practitioners, as everyday musical experiences are increasingly recognized as relevant to music education. Typically, young children's spontaneous singing has been studied in educational contexts where social, or interpersonal, forms of singing dominate musical play. This means self-directed or intra-personal forms of singing have largely been ignored. In this paper I discuss the self-directed singing of 15 three and four year-old children at home, focusing on how they use singing as a tool for self-management. This is one aspect of a qualitative study examining how three and four year-old children use singing in their everyday home lives. Audio data was collected using the LENA™ all-day recording technology. A total of 187 hours of suitable audio recordings were collected from 15 children (7 boys, 8 girls) aged from 3:0 to 4:10 years (average age 3:8). Analysis of the data revealed that young children sing extensively at home and that their self-directed singing is less likely to adhere to culturally meaningful models than their social singing. The children used singing as a tool to manage the self, including to entertain themselves, enhance their experiences, regulate bodily coordination, express emotion, explore self-identity, protect their self-esteem and to support themselves in being alone. Self-directed singing allows young children to explore improvisation in different ways and may assist in the development of independence and self-reliance. These findings suggest that music educators should provide opportunities for solitary music-making in parallel with more socially-oriented music activities and find ways to support the use of singing in self-management.

ISBN 1-876346-65-5 © ICMPC14

Memory ability in children's instrumental musical practice

Larissa Padula Ribeiro da Fonseca[1], Diana Santiago[1], Victoria Williamson[2]
[1]School of Music, Federal University of Bahia, Salvador, Bahia, Brazil
[2]Department of Music, University of Sheffield, Sheffield, Yorkshire, UK

The intricate relationship between music and memory is of interest to multidisciplinary perspectives. In the field of music education and psychology, this theme has been increasingly investigated much more in adults than with children, despite the importance of understanding how memory develops as children train to be musicians. The present research aimed to examine the relationship between the amount and quality of instrumental practice and memory ability in children. The theoretical references were the working memory model (Baddeley & Hitch, 1974) and the practice model (Barry & Hallam, 2002). The participants were 52 children - 6 and 15 years - residents in Salvador and members of the Pedagogical Experimental Orchestra of NEOJIBA - State Centers of Children and Youth Orchestras of Bahia. The research was divided in two stages. The first stage adopted a descriptive approach in order to record the instrumental musical practice of the young musicians and the routine practice of them over more than one year, including systematic observation, field diaries, recordings, and semi structured interviews. Thematic analysis approach was employed to encompass the rich collection of data obtained. The second stage of the study sought to answer questions concerning the musical and non-musical memory ability in the participants and the relationship between their memory abilities and musical practice. Memory tests for digit, rhythm, pitch and timbre sequences, were specially developed and applied. The results of the first stage revealed that most of the participants met the factors proposed by Barry & Hallam (2002). Similarly, most had high performance in the memory tests. In particular the participants presented a higher score for music information, a fact that confirms the first research hypothesis, that the participants' performance would be better in musical memory tests. However, a significant correlation between the two stages of the research, was not found, most likely due to the format of the tests. This difficulty reflects an urgent need to develop new methodological tools (memory tests suitable for children) in order to conduct further research on the subject of musical performance and memory ability in children. Such research could feed directly into music educational practice as well as cognitive memory theory.

Development and first results from the Musical Ear Training Assessment (META)

Anna Wolf[1], Reinhard Kopiez[1]
[1]Hanover University of Music, Drama and Media, Germany

The testing of musicality has been in effect for 100 years. So far, most tests have focused on the aspect of musical talent and applied the paradigm of detecting small discriminations and differences in musical stimuli.

Less research has been conducted on the assessment of musical skills primarily influenced by an individual's practice. Up to the present day, there has been no comprehensive set of tests to objectively measure musical skills, which has rendered it impossible to measure the long-term development of musicians or to conduct studies on the characteristics and constituents of musical skill.

The following study has, for the first time, developed an assessment instrument for analytical hearing following a strict test theoretical validation, resulting in the *Musical Ear Training Assessment* (META). By means of three pre-studies, a developmental study (n=33) and a validation study (n=393), we verified a one-dimensional test model using sophisticated methods of Item Response Theory identifying the best 53 items to measure a person's ear training skills. For better application, two test versions with 10 and 25 items have been compiled (META-10 and META-25).

Aside from psychometric test development, it was possible to investigate a variety of moderator variables assumed to influence the ear training skill. To our surprise, the participants' main instrument did not influence the META score (d=0.10), nor did whether the participant had learnt some method of solmisation (d=0.11). However, the most-played genre (d=0.20) and, unforeseen by us, the participant's gender (d=0.23; males outperforming females) had significant impacts on the test score ($p<.05$).

Future studies will investigate the mentioned effects, especially the hard-to-explain gender effect, and focus on a comprehensive view of the musician's ear, including notation-evoked sound imagery to explore a general model of musical hearing mechanisms, their influences and educational implications.

Innovative approaches to teaching sight-reading in higher education

Katie Zhukov[1]
[1]School of Music, University of Queensland, Brisbane, Queensland, Australia

This paper reports on several research projects that focused on developing innovative approaches to enhancing sight-reading skills in higher education pianists. Zhukov (2006) identified three promising areas of pedagogical applications of sight-reading research: accompanying (Lehmann & Ericsson 1993, 1996; Wristen, 2005), rhythm training (Fourie, 2004; Henry, 2011; Kostka, 2000; Smith, 2009) and understanding of musical styles to improve pattern recognition and prediction skills (Sloboda et al. 1998; Thomson & Lehmann, 2004; Waters, Townsend & Underwood, 1998). The first project evaluated effectiveness of new distinct training strategies (accompanying, rhythm, style) against a control. One hundred participants were tested twice on sight-reading, 10-weeks apart. The MIDI files of playing were analysed by tailor-made software, giving two pitch and two rhythm scores. Mixed-design ANCOVAs administered to test data showed that each treatment group improved in one pitch and one rhythm category, and control group became more accurate in pitch. The findings demonstrated that the new approaches did have an impact on improving sight-reading skills and provided some numerical evidence to the old myth that simply doing more sight-reading could enhance sight-reading (control). In the second project the three training strategies were amalgamated into a hybrid approach, with each week of the 10-week program containing rhythm training, four solo pieces of different styles and collaborative playing items. The course materials were developed in collaboration with piano staff from four institutions who trialled the prototype curriculum with two different student cohorts. Mixed-model ANCOVAs showed a significant improvement in all rhythm and pitch categories in the hybrid-approach group, thus exceeding previously achieved outcomes. The proven effectiveness of the hybrid-training approach led to publication of a higher education textbook on sight-reading.

The Relationship Between Kindergarten Children's Vocal Performance and Recognition of Songs with and without Words

Ana I. Pereira,[*1] Helena Rodrigues,[*2] Patrícia Ávila[#3]

[*] *CESEM, Universidade NOVA de Lisboa (FCSH-UNL), Portugal*

[#] *CIES, University Institute of Lisbon (ISCTE-IUL), Portugal*

[1] anaisabelsaint@gmail.com, [2] helenarodrigues@musicateatral.com, [3] patricia.avila@iscte.pt

ABSTRACT

Background

The ability to recognize songs has been of interest to researchers over the past decades (Racette & Peretz, 2007; Samson & Zatorre, 1991; Serafine, Davidson, Crowder, & Repp, 1986). Evidence on song perception has shown that melody and words are processed with different degrees of integration and separation (Nakada & Abe, 2009; Saito et al., 2012; Sammler et al., 2010). However, there are few studies investigating song recognition among children (Feierabend, Saunders, Holahan, & Getnick, 1998; Morrongiello & Roes, 1990; Rodrigues & Rodrigues, 2010).

Furthermore, how songs are taught both with words and without words, i.e., with a neutral syllable, and how words influence perception of melody and words have not been addressed in depth. Also, results from studies regarding the influence of teaching both types of songs on children's vocal performances were inconclusive and, in some cases, contradictory. For example, some researchers found that children sing better with words (Levinowitz, 1989; Phillips, 1989). In contrast, other investigations did not support these findings and showed no statistical differences between approaches (Jacobi-Karna, 1996; Lange, 2000; Smale, 1988). Within this context, it is unclear if vocal performance of songs with words and without words is related to the way children perceive melody and words.

Aims

The purpose of this study was to investigate the nature of the relationship between children's decisions on recognition tasks based on a song with words and another song without words, both previously taught by rote, and the vocal performance of those songs. The specific goals of this study were (a) to find out how kindergarten children think when judging changes made on melody or words of two familiar songs, (b) to determine whether children sing better if a song is presented with words or without words, and (c) to determine if the means through which children perceive songs is related to its vocal performance.

Method

Fifty-two kindergarten children (4-6 years old) attending a private school in Lisbon and belonging to families with medium/high income levels participated in this two-phase study.

In phase one (instruction phase), the first author taught the participants over a period of 5 weeks during their 30-minute weekly music sessions. Two unfamiliar songs were used: song A with words and song B without words (with the syllable "bá"). Songs were similar in tonality (major), meter (duple), length and range (D_3–A_3). During this procedure, children were taught to associate a different gesture to each song: for song A, pulse was tapped with both hands over the head (gesture 1); for song B, pulse was tapped with one hand on the stomach (gesture 2). The purpose of this activity was to create a tool to use in the test phase.

In phase two (test phase), participants were individually asked to sing both songs and tested on three recognition tasks. The stimulus set was already recorded: i) song A without words (for task 1); ii) song B with words (for task 2); iii) song A with melody of song B (for task 3). Before each listening, the following instruction was provided: "Let's play a game. You are going to listen to me sing. After that, you have to say which gesture you think fits better for the melody you have heard: this one (the researcher exemplified gesture 1) or this one (with gesture 2)." After students chose the gesture, the researcher asked: "Why do you think that gesture fits better?". The average length of an interview was 10 minutes.

Three independent judges—music educators who work regularly with kindergarten children—rated children's performances using a researcher-developed performance rating scale for each song. Each scale comprised tonal and rhythm dimensions with five criteria each.

Results

Children's responses revealed three different behaviors for song recognition tasks: word-based (13 participants), melody-based (11 participants), and both, depending on the task (26 participants). Different expressions were found to explain changes introduced in songs A and B: "it is like," "it reminded me of," "it sounds like," "it is the same," "it makes the same noise," "it is somewhat the same," "it has the same melody," "it has the same words." Also, researchers observed the same behavior with the choice of different gestures.

Cronbach's alpha was used to determine interjudge reliabilities. Coefficients were high, positive, and significant for both rating scales: for song A, α_A= .959, and, for song B, α_B= .932. T-test results for paired samples demonstrated that mean scores for song A were significantly higher (M = 1.75, SD = 2,56) [t = 4.829 (49), ρ < .001]. Also, results showed that eight participants performed better on song B, 29 performed better on song A, and 13 showed no relevant differences between both songs' performance. Results of a one-way ANOVA demonstrated that there was no significant relationship between age and the differences between vocal performances (F= 1.637, ρ= .205), although there was a tendency for the 4-year-old to perform song A better than B (4-year-old, M = 2.69, SD = 0.48; 5-year-old, M = 2.35, SD = 0.81; 6 year-old, M = 2.21, SD = 0.89).

Results showed no significant interaction between preferred criteria for recognition (words, melody, or both) and vocal performances on both songs (F= .575, ρ= .567).

Conclusions

This study demonstrated different ways of perceiving melody and words of a song. Although the results did not reveal a relationship between recognition of songs with and without words and its vocal performance, findings indicate that individual differences should be accounted for among kindergarten children. Therefore, it is important to consider songs without words in classroom activities. This is valid not only for vocal performance benefit, but also for helping students concentrate on the melody and not become distracted by words. Further studies should replicate these procedures with older children and different songs. Also, effects of teaching strategies should be considered in longitudinal studies regarding song perception and production.

ACKNOWLEDGMENT

This research was developed in the framework of CESEM (Sociology and Musical Aesthetics Research Center – UID/EAT-00693/2013) and financial support was provided by FCT (Fundação para a Ciência e Tecnologia) through a PhD grant (PD/BD/114489/2016), co-financed under the European Social Fund (ESF) and the Ministry of Science, Technology and Higher Education (MCTES). The authors would like to thank Daniel Albert for his helpful comments and review of this abstract.

REFERENCES

Feierabend, J., Saunders, T., Holahan, J., & Getnick, P. (1998). Song recognition among preschool-age children: An investigation of words and music. *Journal of Research in Music Education, 46*, 351-359.

Jacobi-Karna, K. (1996). The effects of inclusion of text on the singing accuracy of preschool children. Dissertation Abstracts International, 57(11), 4682A. (UMI No. 9713419)

Lange, D. (2000). The effect of the use of text in music instruction on the tonal aptitude, tonal accuracy, and tonal understanding of kindergarten students. *Dissertation Abstracts International, 60*(10), 3623A. (UMI No. 9948130)

Levinowitz, L. (1989). An investigation of preschool children's comparative capability to sing songs with and without words. *Bulletin of the Council for Research in Music Education, 100*, 14-19.

Morrongiello, B., & Roes, C. (1990). Children's memory for new songs: Integrations or independent storage of words and tunes? *Journal of Experimental Psychology, 50*, 25-38.

Nakada, T., & Abe, J. (2006). The interplay between music and language processing in song recognition. In M. Baroni, A. Addessi, R. Caterina & M. Costa, *Proceedings of the 9th International Conference on Music Perception & Cognition* (ICMPC9). Bologna: Italy.

Phillips, K. (1989). Mary Goetze: Factors affecting accuracy in children's singing [Review of the unpublished doctoral dissertation *Factors affecting accuracy in children's singing*, by M. Goetze]. *Bulletin of the Council for Research in Music Education, 102*, 82-85.

Racette, A., & Peretz, I. (2007). Learning lyrics: To sing or not to sing? *Memory & Cognition, 35*(2), 242-253.

Rodrigues, H., & Rodrigues, P. (2010). Reconhecimento de canções por crianças de quatro a cinco anos de idade: palavra ou melodia? *Revista de Educação Musical, 134*, 11-22.

Saito, Y., Ishii, K., Sakuma, N., Kawasaki, K., Oda, K., & Mizusawa, H. (2012). Neural substrates for semantic memory of familiar songs: Is there an interface between lyrics and melodies? *PLoS ONE 7* (9), e46354.

Sammler, D., Baird, A., Valabrègue, R., Clément, S., Dupont, S., & Samson, S. (2010). The relationship of lyrics and tune in the processing of unfamiliar songs: A functional magnetic resonance adaptation study. *The Journal of Neuroscience, 30*(10), 3572-3578.

Samson, S., & Zatorre, R. (1991). Recognition memory for text and melody for songs after unilateral temporal lobe lesion: Evidence for dual encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(4), 793-804.

Serafine, M. L., Davidson, J., Crowder, R., & Repp, B. (1986). On the nature of melody-text integration in memory for songs. *Journal of Memory and Language, 25*(2), 123-135.

Smale, M. (1988). An investigation of pitch accuracy of four- and five-year-old singers. *Dissertation Abstracts International, 48*(08), 2013A. (UMI No. 8723851)

The development of an assessment for notation-evoked sound imagery (NESI)

Anna Wolf[1], Reinhard Kopiez[1], Friedrich Platz[2] and Luis Alfonso Estrada Rodriguez[3]
[1]Hanover University of Music, Drama and Media, Germany
[2]State University of Music and Performance Stuttgart, Germany
[3]Universidad Nacional Autónoma de México, Mexico

The successful transformation of music from the written score to being played with accurate movements and to its being heard lies at the heart of skilled musicianship. To enable future research on the development of musicians, the processes of this transformation have to be investigated more thoroughly. The aim of this study is to further develop a rapid assessment instrument to observe the transformation from the written score to the imagery of sounding music: namely, the concept of notation-evoked sound imagery (NESI), which follows the notion of *notational audiation*.

Referring to Brodsky et al. (2008), we have developed and validated n=26 sets of stimuli consisting of a theme melody, a correct variation and an incorrect variation ("lure") of this particular theme. In this procedure, participants (music students and musicians, N=43) read the theme, try to imagine the sound, hear a correct or incorrect variation and decide whether the heard variation is based on the seen theme. These stimuli have been examined in Pre-study 1 and, by the means of signal detection theory, analysed with regards to their item characteristics (Sensitivity/Easiness d': M=1.30, SD=0.84. Bias c: M=0.07, SD=0.45). In the upcoming Pre-study 2, both easier and more difficult items are being generated and will be tested in spring 2016. Several control variables will be included based on the preliminary results from the first study.

On the basis of these pretested items, a test instrument will be developed to measure the extent to which musicians are able to imagine the sound represented by a written score (inner hearing). As many musical skills (e.g., sight-reading, playing by ear…) rely on the fast and accurate switching between sounding music and the written score, assessments will both facilitate research in analytical hearing and mental representations of music and uncover the role of the musical ear for musicians of all skill levels.

Cross-species studies of music cognition

Organizer: Aniruddh D. Patel[1]
Discussant: Henkjan Honing[2]
[1]Dept. of Psychology, Tufts University, Medford, MA, USA
[2]Music Cognition Group, University of Amsterdam, The Netherlands

Recent years have seen increasing interest in empirical cross-species research on music cognition, as a way of addressing evolutionary questions about music. This symposium brings together researchers who are doing comparative studies of music cognition in chimpanzees, sea lions, and songbirds to address questions about the origins of human rhythmic and melodic processing capacities. The symposium will include research based on a variety of empirical methods (behavioral and neural) and will present new empirical research relevant to a range of theoretical issues, including the origins of human beat perception and synchronization and cross-species similarities and differences in the perceptual cues used to recognize melodic patterns. The four speakers and the discussant come from labs in the US, Japan, and Europe, reflecting international research on music cognition. All have published research in the area of cross-species music cognition, and all will include new data in their presentations.

Rhythmic engagement with complex beat in chimpanzees

Yuko Hattori[1,2], Masaki Tomonaga[2], Tetsuro Matsuzawa[2]
[1]Wildlife Research Center, Kyoto University, Kyoto, Japan
[2]Primate Research Institute, Kyoto University, Inuyama, Aichi, Japan

One of the interesting effects of music is that it induces our movement even when simply listening to it. Although recent studies report that non-human apes (and several other animals) also show spontaneous entrainment to auditory rhythms, still it is not fully understood how we acquired such embodied music cognition in the course of human evolution. Here we report spontaneous rhythmic engagement with complex auditory beat in chimpanzees (*Pan troglodytes*). At the experimental booth, we observed that some chimpanzees spontaneously began to coordinate their body movements in response to auditory beats. So, we investigated systematically their responses to different auditory beats by video analysis. The characteristics of rhythmic movements (e.g., rocking, head bobbing) observed were similar to their display they show in daily life and this suggests that movements induced by the beat in the chimpanzees were in the range of their natural behavioral repertoire. We also found more rhythmic engagement to complex beats than to simple beats and stronger effects in male chimpanzees than female chimpanzees. The finding suggests that a predisposition for rhythmic movement in response to music already existed in the common ancestor of chimpanzees and humans, approximately 6 million years ago.

Modeling beat keeping and mapping auditory and motor brain networks in sea lions: Implications for comparative hypotheses of sensorimotor synchronization

Peter F. Cook[1], Andrew Rouse[2], Edward Large[3], Gregory Berns[1], Margaret Wilson[4] and Colleen Reichmuth[2]
[1]Department of Psychology, Emory University, Atlanta, Georgia, USA
[2]Ocean Sciences, University of California Santa Cruz, USA
[3]Department of Psychology, University of Connecticut, Storrs, USA
[4]Department of Psychology, University of California Santa Cruz, USA

Recent comparative findings in birds, primates, and a sea lion have overturned the long-held belief in psychology that beat keeping is unique to humans. It remains to be seen what the limits on beat keeping behavior are in non-human animals, and to uncover the neural underpinnings for flexible rhythmic behavior. We exposed an experienced sea lion beat keeper to multiple tempo and phase perturbations in rhythmic stimuli and modeled her response recovery. Consistent with prior findings in humans, she showed a rapid return to stable phase following phase perturbation and more gradual return following tempo perturbation. These findings suggest the sea lion relies on similar neural mechanisms as do humans, which can be accurately modeled as processes of neural co-oscillation. In parallel work, we used diffusion tensor imaging to map auditory, motor, and vocal production pathways in sea lion brains obtained opportunistically from animals euthanized in a veterinary rehabilitation facility. Tracts were compared to other pinniped species and primates. Based on these and other recent findings, we suggest that beat keeping capability is rooted in basic and widely shared neural processes underlying sensorimotor synchronization. We further suggest that the capability may be enhanced or restricted by a wide range of biological and environmental factors, including faculty for precise motor control, attention, motivation, and social interaction.

Do other animals perceive cross-cultural features of music?

Marisa Hoeschele[1], Daniel L. Bowling[1], Lauren M. Guillette[2], Allison H. Hahn[3], Bernhard Wagner[1], and Christopher B. Sturdy[3]
[1]Department of Cognitive Biology, University of Vienna, Vienna, Austria
[2]School of Biology, St. Andrews University, St. Andrews, UK
[3]Department of Psychology, University of Alberta, Edmonton, Alberta, Canada

Music is found in all human cultures, which suggests that music is not a purely cultural phenomenon, but has roots in our biology. Not only that, but some aspects of music tend to be found cross-culturally, suggesting that they are fundamental to the phenomenon of human music. Are these fundamental features of human musical systems found in other species? If they are, we can begin to understand the evolutionary origins of these features by studying what other animals with these musical abilities have in common with each other and also us. Here we show two examples of methods we are using to understand whether other animals perceive music-relevant sounds the way we do. In both cases, we conducted experiments with humans first without any explicit verbal instructions and then conducted the same experiments with other species to see if they would respond the same way. In the first task, we demonstrated octave equivalence in humans using a paradigm where humans had to respond to some sounds and not to others based on trial-and-error feedback. In the second task, we demonstrated that humans prefer to listen to rhythmic over arrhythmic sounds by placing them in an arena with two rooms connected by an opening, where each room was playing a different soundtrack, and observing where they spent the majority of their time. We have begun to use these simple tasks to test other species. We found that a songbird species that is known for excellent pitch abilities failed to show octave equivalence in the same task that humans completed. We are currently studying how budgerigars, a parrot with excellent pitch, rhythm, and mimicry abilities, fares at both of these tasks. We also plan on studying mammalian species: while some avian species share rare potentially music-relevant abilities with humans, such as vocal mimicry and rhythmic entrainment (synchronizing to a beat), other mammals may share music-relevant genes with humans because they are more closely-related to us than birds.

Do songbirds recognize melody on the basis of absolute pitch? Challenges to a prevailing view.

Aniruddh D. Patel[1], Micah R. Bregman[2], Timothy Q. Gentner[2]
[1]Dept. of Psychology, Tufts University, Medford, MA, USA
[2]Dept. of Psychology, University of California, San Diego, La Jolla, CA, USA

Humans easily recognize 'transposed' musical melodies shifted up or down in log frequency. Surprisingly, songbirds seem to lack this capacity, although they can learn to recognize human melodies and use complex acoustic sequences for communication. Two decades of research have led to the widespread belief that songbirds, unlike humans, are strongly biased to use absolute pitch (AP) in melody recognition. This work relies almost exclusive on acoustically simple stimuli that may belie sensitivities to more complex spectral features. Here, we investigate melody recognition in a species of songbird, the European Starling (*Sturnus vulgaris*), using tone sequences that vary in both pitch and timbre. We find that small manipulations altering either pitch or timbre independently can drive melody recognition to chance, suggesting that both percepts are poor descriptors of the perceptual cues used by birds for this task. Instead we show that melody recognition can generalize even in the absence of pitch, as long as the spectral shapes of the constituent tones are preserved. These results challenge conventional views regarding the use of pitch cues in non-human auditory sequence recognition.

An Empirical Study of Dogs Howling to Music

Justin Yuan,*[1] Jason C. Rosenberg,#[2] Aniruddh D. Patel^[3]

*Department of Psychiatry, University of California San Francisco, USA
#Center for New Music, USA
^Department of Psychology, Tufts University, USA
[1]justin.yuan@ucsf.edu, [2]jason@jasonrosenberg.org, [3]a.patel@tufts.edu

ABSTRACT

A hallmark of human singing is that individuals coordinate their pitch patterns in relation to pitch patterns simultaneously produced by other voices or instruments. Is this ability uniquely human among mammals? One way to examine this issue is to examine the phenomenon of dog howling to music. If a dog howls at specific pitches when hearing a musical piece, one can test whether the dog shifts these pitches when hearing a pitch-shifted version of the piece. If so, this would suggest that the pitch of the howl is being regulated in relation to the pitches in the musical stimulus. As a first step in this line of research, we have quantitatively studied if dogs hit specific pitches when howling to a musical piece. We collected videos of 10 dogs howling to the theme of the TV show "Law and Order". We extracted the pitch contours of each dog's howls to determine if there were segments during which the dog held relatively level pitches. We found that 9 dogs had stable segments when howling, defined as a segment of at least 500ms with no fundamental frequency values more than 5% above or below the mean fundamental frequency of that segment. Focusing on dogs with at least 6 such segments (n = 7 dogs), we found that across dogs, the mean frequency of stable segments ranged from 436 Hz to 1063 Hz, and within a given dog, the standard deviation between the mean frequencies of stable segments ranged from 1.08 semitones to 8.1 semitones). For 3 of the 7 dogs, the standard deviation of the stable segments was less than 2 semitones. This shows that at least for some dogs, their stable segments when howling to music are restrained to a fairly narrow range of pitches. Such dogs would be ideal to study by playing them pitch-shifted versions of howl-inducing songs, to determine if they shift their vocal pitches in response.

I. INTRODUCTION

Singing in groups is a widespread feature of human musicality, and involves coordinating one's own vocal pitch pattern with structured pitch patterns simultaneously produced by others. This type of vocal pitch coordination (which can also occur in chanting) is rarely observed in other human behaviors. Ordinary speech, for example, involves turn taking, and thus does not require coordinating vocal pitch production with other simultaneously-sounded pitch patterns.

Are humans unique among mammals in their ability to regulate the pitch of the voice in relation to other concurrently-heard pitch patterns? If so, this might suggest that this ability reflects a neural specialization that arose to support group musical behavior, e.g., because it allows voices to blend into a coherent whole, thus enhancing the social-bonding function of music (cf. Tarr et al., 2014). On the other hand, if there are non-human mammals with this capacity, this would suggest it did not first arise as a specialization for musical behavior. Furthermore, determining which other mammals have this capacity would help address questions about the evolutionary history and original functions of this ability.

In addressing this issue, it is important to distinguish the ability in question (voluntary coordination of vocal fundamental frequency in relation to simultaneously-sounded pitch patterns) from vocal learning. Vocal learning involves learning to imitate an auditory model, and this model is typically not heard at the same time that one is vocalizing. For example, young songbirds learn their songs from adult birds, but not by 'singing along' with adults and coordinating vocalizations with them. Rather, periods of listening and singing are typically temporally separated in time.

Recently, Pisanski et al. (2016) have argued that non-human apes and some other mammals may have a degree of voluntary control over their fundamental frequency (F0) patterns, which could be used for communicative purposes. As noted by these authors, the qualification of 'voluntary' is important, because F0 can change in mammalian vocalization as a function of arousal state. Such involuntary regulation of F0 is not relevant to the current discussion, which is concerned with the purposeful coordination of one's own F0 patterns with others. Is there any evidence that great apes (our closest genetic relatives) voluntarily regulate their F0 patterns in relation to each other when vocalizing simultaneously? Chimpanzees are known to engage in bouts of group vocalization. For example, the 'carnival display' is a vocal and motor display of excitement made by foraging males who discover a ripe fruiting tree, which serves to attract other males and females to the site (Merker et al., 2009). It is not known, however, if F0 patterns are purposefully coordinated as part of this display. A priori, this seems unlikely, as Merker et al. suggests that the display shows little sign of inter-individual coordination.

Wolves and their descendants (domestic dogs) may prove a more tractable group of animals than apes for studying the issue at hand. These canids engage in a well-known vocal behavior, howling, which often involves group vocalization. Like human singing, howling involves sustained vocalizations with perceptually salient pitch patterns, which can include significant F0 modulation. Theberge and Falls (1967) studied the howls of timber wolves raised in captivity and described the howl "as a continuous sound, from 0.5 to 11.0 seconds long, with a fundamental pitch from 150 to 780 Hz, and up to 12 harmonically related overtones" (quoted in Harrington & Mech, 1978, p. 116). Howling is a long-distance signal that appears to serve social functions, including assembling separated pack members and serving as a territorial advertisement aimed at other packs (Harrington & Merch, 1978). An acoustic study by Tooze et al. (1990) suggested that individual timber wolves produce distinctive howls, distinguished by mean F0 and F0 variability. A perceptual study by Palacios et al. (2015) suggests that Iberian wolves may be able to discriminate between individuals on the basis of their howls, and a behavioral and physiological study of timber wolves by Mazzini et al. (2013) suggests that howling is not simply a byproduct of increased stress, and may be under flexible control of the signaler (cf. Fitch, 2000).

If individual wolves regulate the F0 of their howls in relation to the howls made by others, this would show that the capacity for regulating one's own pitch in relation to other, simultaneously heard pitches is not uniquely human. An acoustic study by Filibeck et al. (1982) examined chorus howls made by four canids howling simultaneously (two captive wolves, a dog-wolf hybrid, and dog) and found that the F0s of the animals differed from each other by at least 15 Hz. This finding accords with the speculation that chorus howling involves purposeful frequency separation between the howls of individual wolves, which may enhance the apparent size of the pack to other canid groups listening to such howling (Lopez, 1978). While this idea is speculative, it raises the interesting possibility that when a wolf (or dog) howls along with others, it purposefully 'detunes' its F0 from the other pitches that it hears.

We propose that a suitable model system for addressing the voluntary vocal pitch flexibility of canids is provided by the phenomenon of domestic dogs (*Canis lupus familiaris*) howling to music. This is a frequently-observed behavior, as evidenced by many videos of the behavior posted on public video-hosting websites (such as YouTube). Howling to music is easy to distinguish from other vocalizations dogs might make while music is playing (such as barking, growling, or whining) because of the distinctive acoustic structure of the dog howl (Taylor et al. 2014) and the characteristic posture associated with it (muzzle pointing up). Dog howling to music is likely elicited by pitch patterns in the music that resemble the acoustics of howls, i.e., sustained pitches with salient periods of frequency modulation (hence the canine tendency to howl to non-musical sounds with these qualities, e.g., ambulance sirens or fire trucks).

Anecdotal reports (e.g., from dog owners who post videos online) suggest that some dogs howl fairly reliably when they hear certain pieces of music. If this is the case, it provides an opportunity to determine whether dogs change the pitch of their howls in relation to simultaneously heard pitch patterns. Specifically, one could analyze the pitches of a dog's howl to a particular musical piece and to versions of the piece transposed up or down by a few semitones in pitch. If the dog changes the pitch of its howl in response to such transpositions, this would provide evidence that it is attempting to coordinate its own pitch pattern with the pitch patterns it hears (e.g., to purposefully detune its howl from those patterns). Importantly, research on dog auditory perception suggests that they have the hearing range and frequency discrimination abilities to be able to discriminate between versions of a musical piece transposed up or down by a few semitones. The dog hearing range spans from about 67 Hz – 45 kHz at 60 dB SPL (Heffner, 1983), and dogs can discriminate pitch differences of 1 semitone or less within musically-relevant pitch ranges (Anrep, 1920).

As a first step toward the proposed transposition experiment, we have conducted an empirical study of dogs howling to music. Our question was how often dogs hit stable pitches when howling to music. By 'stable pitches' we do not mean pitches corresponding to Western musical scales, but sections of pitch contour with relatively little frequency variation. The presence of stable pitches allows one to study the relation of howl pitches to the pitches in the musical structure. Stable pitches are also ideal from the standpoint of the proposed transposition study, because one can measure the mean F0 of stable pitches taken from dogs howling to a piece of music in its original and transposed versions. A significant difference in the mean F0 of such pitches before and after transposition would suggest voluntary regulation of vocal pitch in relation to heard pitch patterns.

The current study did not perform the transposition experiment, which we hope to conduct in the future. Here we focus on determining how often stable pitches occur when dogs howl to music, and how much variability there is between the mean frequencies of such stable pitches within a given dog's howling to music.

II. METHODS

A. Sample

29 videos of dogs howling to a range of musical stimuli were gathered using YouTube and Google search engines. The most common music within these videos was the 2010 theme to the television show Law & Order (n=10), a piece of about 40 seconds duration (tempo = 118 BPM) performed by electric guitar, keyboard, timpani, and clarinet. Videos of dogs reacting to this theme were selected for further analysis (these dogs spanned a range of different breeds and sizes). The audio content of the videos and the 2010 Law & Order theme were extracted (WAV format, mono, 16-bit, 48kHz) using an online audio extraction tool (http://www.onlinevideoconverter.com).

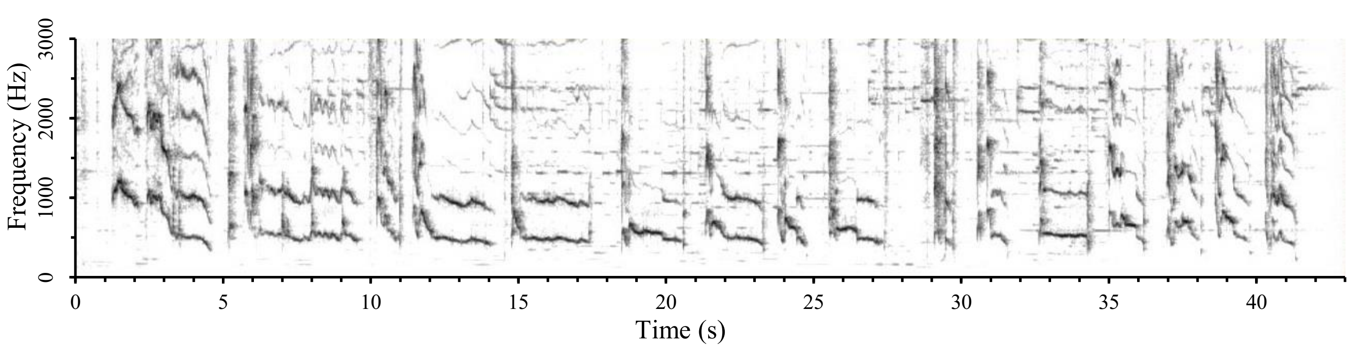


Figure 1. A spectrogram of a dog howling in response to the musical stimulus.

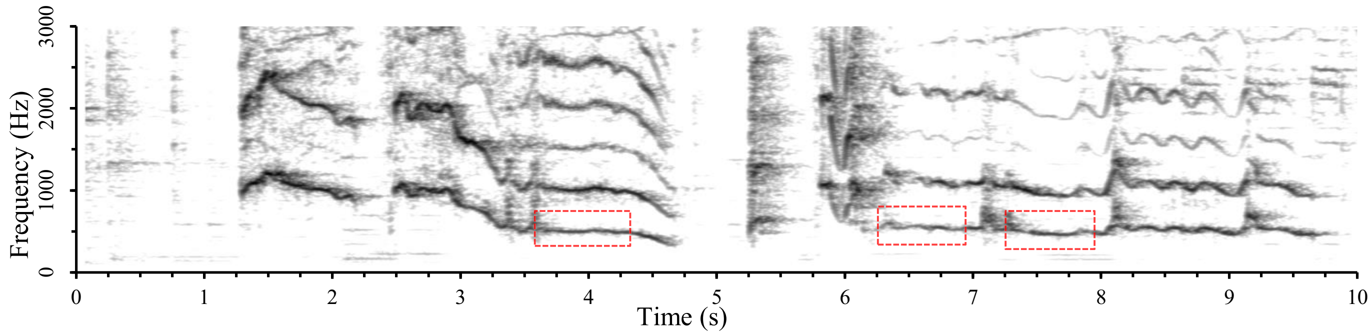


Figure 2. An enlarged view of the first 10 seconds of Figure 1. Red boxes outline the first three stable howl segments.

B. Howl Quantification

Howls were acoustically analysed using spectrograms made with Praat (Boersma, 2002). Howls were identified using sonic and visual characteristics noted in Tembrock's (1967) classification of canid vocalizations. Pitch (Hz) and time (s) data were measured for each howl. Praat default pitch and spectrogram settings were used, with two deviations (Pitch Range = 50.0 – 2500.0 Hz and Very Accurate = On). Stable howl segments were identified from pitch contour data imported into Excel 2016, using the following criteria:

1) Pitch. The maximum and minimum fundamental frequency of the segment was within ±5% of that segment's average F0.

2) Duration. The segment was at least 0.5s in duration.

One video had no segments that met the criteria and could not be analyzed further, leaving 9 audio tracks for further analysis. An example spectrogram from our study is shown in Figure 1. The F0 and first few harmonics of the howls are clearly visible in this figure, and are much more prominent than the background music, which is barely visible on the spectrogram. Figure 2 enlarges the first 10 seconds of the spectrogram shown in Figure 1, and red boxes outline stable howl segments identified using the above criteria.

C. **Alignment**

The 2010 Law & Order Theme was transcribed into music notation, and a 4-part MIDI version of the theme was created in Ableton Live 9. Extracted audio from dog howling videos were then aligned to this reference, using the music in each video for alignment with the original theme song. In two videos, the tempo of the music was slower than the tempo of the original music, suggesting that the uploaded video had undergone temporal dilation (in one video, the tempo of the music was 112 BPM and in another, 87 BPM). Time warping in Ableton Live 9 was used to bring these audio files up to the original tempo of 118 BPM (without changing pitch) and align them to the original theme song.

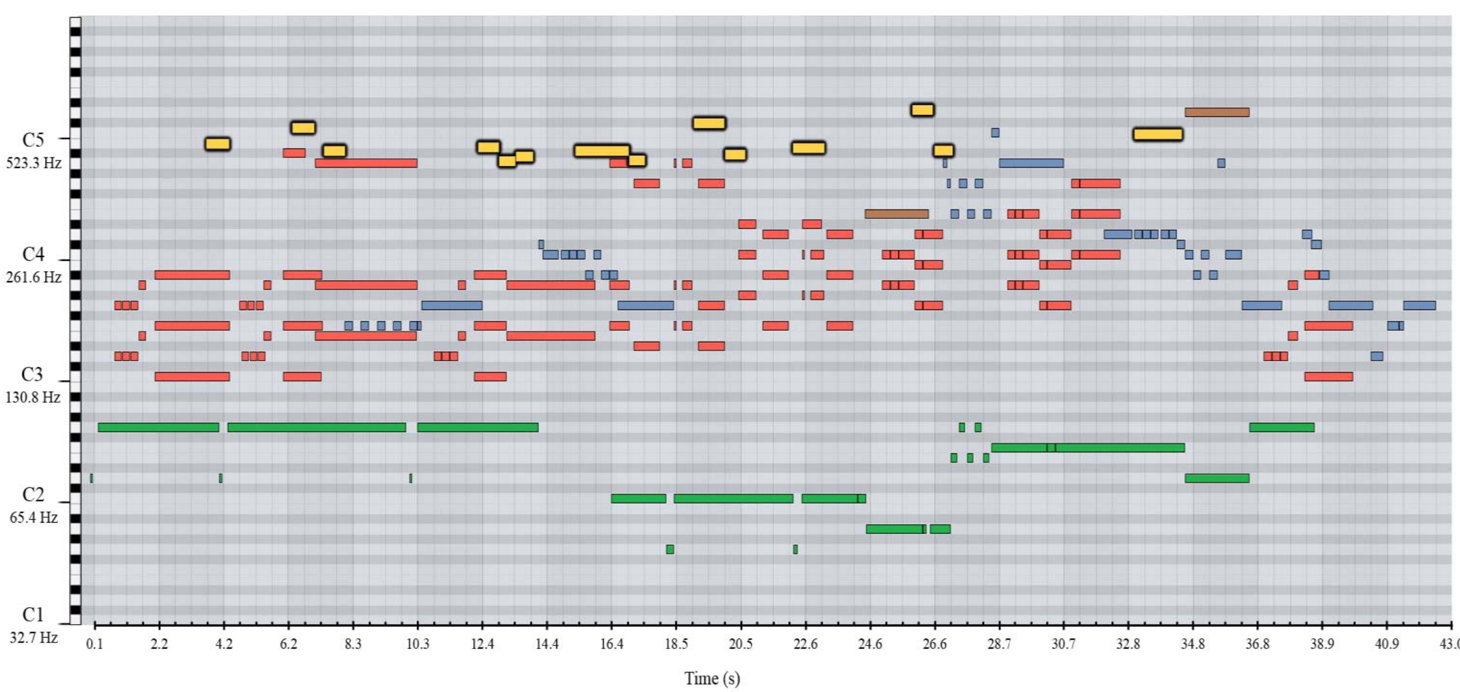


Figure 3. A piano roll visualization of stable segments of dog's howls relative to pitches in the musical stimulus (analysis based on the howls in Figure 1). Darker colors (blue, red, green, and brown) represent instruments of the Law & Order theme (guitar, keyboard, timpani, and clarinet, respectively). Yellow indicates the average pitch of stable howl segments. These segments have not been rounded to Western chromatic pitches.

D. Piano Roll Plot

For each dog, stable howl segments were plotted as pitch (cents) vs. time (s) in Excel 2016. Frequency values were converted from Hz to cents with the following:

$$c = 1200\log_2(\frac{f_1}{f_2}) \qquad (1)$$

where c = cents, f_1 = howl frequency, and f_2 = 440Hz. The resultant plot was then overlaid on top of an image of the MIDI reference viewed in Ableton to create the piano roll visualization. An example of this visualization is shown in Figure 3. Yellow segments indicate average pitch values from stable howl segments. These values are not rounded to Western chromatic pitches: they can fall in between such pitches. Such figures were then used to identify possible relationships between stable howl segments and musical elements of the Law & Order theme. The majority of notes from the clarinet were omitted from the piano roll plot due to their short duration and quickly-changing pitches. (Specifically, the piano roll plot focused on pitches that were either sustained for 1 second or quickly repeated over at least 1 second, since we reasoned these would be the most perceptually relevant pitches for dogs attempting to coordinate their howls with the music.)

III. RESULTS AND DISCUSSION

9 dogs had stable pitch contour segments when howling, when using the criteria outlined in the Methods to identify such segments. We focused further analysis on dogs who produced at least 6 such stable segments over the course of their howling to the musical piece (n=7 dogs). There were 74 such segments across these 7 dogs, with a mean duration of 820ms (SD=352ms). Across dogs, the average frequency of stable segments ranged from 436 Hz to 1063 Hz, and within a given dog, the standard deviation between the mean frequencies of stable pitches ranged from 33 Hz to 213 Hz (1.08 semitones to 8.1 semitones). For 3 of the 7 dogs, the standard deviation of their level segments was less than 2 semitones, indicating that such segments fell in a fairly narrow pitch range (i.e., less than a major second, in musical terms).

There were no discernible relationships between the timing of the stable segments the rhythmic characteristics of the Law & Order theme music. In terms of pitch, we observed no obvious relationship between the pitches of the stable segments and pitches of the music. We feel that a larger amount of data are necessary to empirically study whether dogs might be aligning or 'detuning' their howl pitches from the pitches in the music. We can, however, conclude that there are dogs who produce multiple stable pitch segments when howling to music, and that for some dogs, these segments fall in fairly narrow pitch range. This suggests that the proposed transposition study (outlined in the Introduction) is feasible and worth pursuing. The results of such a study would help determine whether humans are unique in their ability to coordinate their vocal pitch patterns with structured pitch patterns simultaneously being produced by others.

REFERENCES

Anrep, G.V. (1920). Pitch discrimination in the dog. *The Journal of Physiology, 53,* 367-385.

Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glot international*, *5*, 341-345.

Filibeck, U., Nicoli, M., Rossi, P., & Boscagli, G. (1982). Detection by frequency analyzer of individual wolves howling in a chorus: A preliminary report. *Italian Journal of Zoology*, *49*, 151-154.

Fitch, W.T. (2000). The phonetic potential of non-human vocal tracts: comparative cineradiographic observations of vocalizing animals. *Phonetica, 57,* 205-218.

Harrington, F. H., & Mech, L. D. (1978). Wolf vocalization. In R.L. Hall & H.S. Sharp (Eds.), *Wolf and Man: Evolution in Parallel* (pp. 109-132). New York: Academic Press.

Heffner, H.E. (1983). Hearing in large and small dogs: Absolute thresholds and size of the tympanic membrane. *Behavioral Neuroscience, 97,* 310-318.

Lopez, B. (1978). *Of wolves and men*. New York: Simon and Schuster.

Mazzini, F., Townsend, S.W., Viranyi, Z., and Range, F. (2013). Wolf howling is mediated by relationship quality rather than underlying emotional stress. *Current Biology, 23,* 1677-1680.

Merker, B. H., Madison, G. S., & Eckerdal, P. (2009). On the role and origin of isochrony in human rhythmic entrainment. *Cortex*, *45*, 4-17.

Palacios, V., Font, E., Márquez, R., & Carazo, P. (2015). Recognition of familiarity on the basis of howls: a playback experiment in a captive group of wolves. *Behaviour*, *152*, 593-614.

Pisanski, K., Cartei, V., McGettigan, C., Raine, J., & Reby, D. (2016). Voice modulation: a window into the origins of human vocal control? *Trends in Cognitive Sciences*, *20*, 304-318.

Tarr, B., Launay, J., & Dunbar, R.I. (2014). Music and social bonding: "self-other" merging and neurohormonal mechanisms. *Frontiers in Psychology, 5:*1096 *doi: 10.3``389/fpsyg.2014.01096*

Taylor, M., Ratcliffe, V.F., McComb, K., & Redy D. (2014). Auditory communication in domestic dogs: Vocal signaling in the extended social environment of a companion animal. In J. Kaminsi & S. Marshall-Pescini (Eds.), *The Social Dog: Behavior and Cognition* (pp. 131-163).

Tembrock, G. (1976). Canid vocalizations. *Behavioural Processes*, *1*, 57-75.

Theberge, J. B., & Falls, J. B. (1967). Howling as a means of communication in timber wolves. *American Zoologist*, *7*, 331-338.

Tooze, Z. J., Harrington, F. H., & Fentress, J. C. (1990). Individually distinct vocalizations in timber wolves, *Canis lupus. Animal Behaviour*, *40*, 723-730.

Out of sight, into mind

Jonathan Govias[1]
[1]University of North Carolina Charlotte, Charlotte, NC, USA

Synchrony, as applied to music, is typically defined as the capacity of musicians to maintain rhythmic precision in performance. Without exception, the current research on the topic of musical synchrony accepts the need for visual contact between all performers involved. The presentation proposed herein challenges that base assumption. Under the premise that music is an auditory art, the project investigates the hypothesis that all acts of synchrony, be they: the simultaneous onset and release of notes; the initial establishment of tempo; deliberate acts of tempo modifications and fermatas or other unmeasured musical events; can be executed without visual communication but through non-verbal auditory cues alone. To test this hypothesis, a quartet was assembled ad hoc to rehearse and perform the Dvořák Bagatelles Op. 47. After 2 rehearsals in conventional inward-facing configuration, the remaining rehearsals and subsequent performances were executed in a "back-to-back" configuration with all musicians facing outwards, deprived of any visual contact with each other. Approximately 18 hours of rehearsal and performance were audio and video recorded for analysis. Audio from an "unsighted" performance was then contrasted with audio from a "sighted" performance of the work by the same musicians, with both samples, unidentified by performance process, evaluated by qualified external referees. As an outcome, the "unsighted" quartet was able to achieve levels of synchrony indistinguishable from or occasionally superior to the same quartet in an inward-facing configuration. The experiment also revealed the critical role physical motion plays as a means of achieving synchrony for musicians even when in the "unsighted" configuration. The results challenge the perceived need for the traditional directive or visual style of ensemble leadership, and offer music educators new possibilities and tools for facilitating group musical action for both musical and social outcomes.

Movement effector training and musical expertise heighten temporal detection abilities

Fiona Manning[1], Anna Siminoski[1] and Michael Schutz[2,1]
[1]Department of Psychology, Neuroscience and Behaviour, McMaster University, Canada
[2]School of the Arts, McMaster University, Canada

Sensorimotor integration is typically studied using finger tapping studies. However, synchronizing movements with an external pacing signal using a drumstick is both more synchronous and less variable that is finger tapping, particularly for percussionists. This raises questions about the degree to which musicians rely on trained movements for internal timing representations of external auditory information. In the present set of studies, musicians of different instrumental groups completed a timing detection task wherein each group was required to identify temporal deviations after synchronizing with motor effectors consistent with their training, with finger tapping or tapping with a drumstick. We found that while musicians synchronized with a large degree of consistency using training-specific movement, they were most consistent while tapping with a drumstick. The consistency of these movements extended to temporal detection abilities, where musicians were more accurate in detecting temporal deviations in an external stimulus when trained movements accompanied the stimulus, and were most accurate with stick tapping movements. We suggest that movement offering more sensory feedback leads to more consistent internal timing representation, regardless of training. However, training consistent movements appear to also aid with timing detection abilities. These findings expand current thinking surrounding motor training and musical expertise, and how the consistency of movement timing can improve predictive timing abilities.

Vocal learning – the acquisition of linguistic and musical generative grammars

Stefanie Stadler Elmer

The Schwyz University of Teacher Education & University of Zurich, Switzerland
stefanie.stadlerelmer@uzh.ch

Abstract

This paper outlines a theoretical framework for reconstructing functional and structural dimensions of vocal learning and illustrates with micro-genetic analyses how children organise and change their song singing. Vocal learning means using the capacity to hear sounds produced by outside or inside sources, to adjust vocally to the sensed sounds, to use feedback to monitor the sensorimotor matching process, and to build up neural pathways and memory traces. As a unique human characteristic, vocal learning leads to singing and speaking. Based on physical preconditions, vocal teaching and learning – the intergenerational transmission of culture – is motivated by the human basic need to belong. The desire for interpersonal attachments is physically grounded in the propensity to entrain or to synchronize (vocal-) motor movements with others and within oneself. Self- and interactive synchronization, together with intuitive parenting, are basic prerequisites for vocal learning and communication. No other species has the capacity to construct a similar complexity of spoken language systems as well as technics and rules of vocal musical production. Vocal products – speech and songs – are governed by generative rules, i.e. the fact that language and music generate infinite pattern diversity by finite means. In songs, linguistic and musical rules are combined. As a cultural ritual, songs are devices used to share and to modulate affective states. A grammar of children's songs is outlined that allows conceptualizing children's strategies in making up a new song. Micro-genetic analyses consist of 1) acoustical analysis and 2) of visualizing the vocal production as a configuration of syllables (vowel-consonants), their timing (onset, duration), accents (stressed, unstressed), and pitches. This method allows reconstructing step-by-step how a child applies generative rules to create a new song, and thereby strive for well-formedness or a coherent Gestalt.

I. Introduction

Many disciplines are interested in the question on how new qualities or changes emerge in human behavior. The use of the voice to signal affective states and to communicate with others is a common phenomenon in the animal world. Yet, it seems a specific human capacity to cultivate vocal expressions in form of speaking, reciting, and singing and thereby make use of our ancestors' evolutionary cultural achievement of having constructed language and music systems. Theorizing about the ontogeny of speech and song and the emergence of language and music needs specifying potential necessary conditions such as the early physical constraints, the nature of social care and communication, the structures of speaking and singing, their functions and cultural contexts, for investigating the varieties of developmental conditions and trajectories, in order to understand the rapid vocal learning in early childhood and the general construction of complex systems such as language and music. Here, the aim is to present a conceptual framework to describe and analyze changes in children's vocal utterances whilst they intend to acquire or create a new song.

II. Vocal learning

Vocal adaption to cultural specific features has been evidenced in French and German born neonates and their melodic contour of their crying (Mampe et al.). From early on infants begin to adapt their vocal expression to the speaking and singing models provided by their caretakers. Vocal learning can be defined as the capacity to adjust the own vocal expression to sounds heard in the environment, typically to the song and speech models. Necessary conditions are i) intact sensorimotor functioning of hearing, the voice, and neural correlates, and ii) the capacity to co-ordinate sound perception with the voice, to use feedback, to construct representations, and to monitor the production; iii) the motivation to engage in this process. Figure 1 illustrates main aspects of vocal learning.

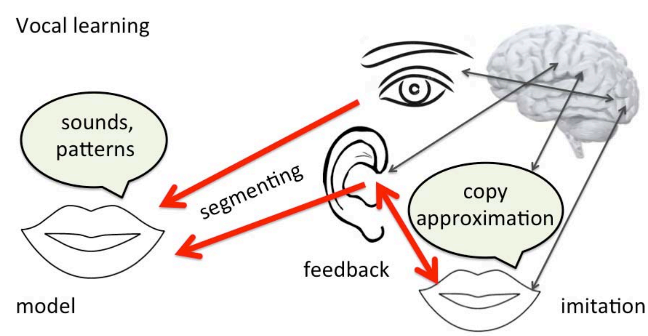


Figure 1: A graphic representation of vocal learning (© Stadler Elmer, 2015).

Furthermore, early vocal learning requires a specific quality of human interaction and care that is described by Papousek and Papousek (1987) as intuitive parenting.

III. Motivation to engage in vocal learning

The participants in the intergenerational transmission of cultural achievements such as songs and speech – the youngsters and parents – seem to be highly motivated to share the affective experiences to play and vocalize together long before serious conversations or artistic performances are possible. What are the functions of engaging in vocal play, early singing, and proto-conversations?

Baumeister and Leary (1995) would explain the vocal teaching and learning, joint vocal play, and the sharing of similar experiences as being motivated by the human basic need to belong. In this view, intergenerational transmission of culture – and music and language as integral parts – is a

collective practice to build, maintain, and to cultivate affective bonds between human beings and to satisfy our need to feel socially included. Dissanayake (e.g., 2000, 2008, 2011) analysed in detail early mother-infant interactions, and she theorized about caretakers' proto-aesthetic means to regulate affective and summarized them as repeating, simplifying (or formalizing), exaggerating, elaborating (or varying), and manipulating expectancies (or surprising). The use of these affect regulatory means are precursors of artistic experiences, since they allow to make an event turn into something special, and act Dissanayake calls *artifying.*
Merker (2009) proposes to conceive of human culture as governed by rituals that focus on performing procedures, whereas animal culture is restricted to instrumental usage of tools and communication. He assumes that language and rituals are specific to humans. Moreover, he introduced the idea that music – like language – is a generative system, i.e., the fact that music generates infinite pattern diversity by finite means (Merker, 2002). Hence, vocal products – speech and songs – are governed by generative rules, and in song singing linguistic and musical rules are combined.
The engagement in vocal learning is driven by the infant's basic affective needs such as care, cognitive stimulation, and feelings to belong; and vocal teaching is driven by caretakers' intuition and intention to transmit cultural achievements. Collectively, every culture provides traditional practices to regulate affects by ritual behavior – for instances song singing – that help to make the future somehow less uncertain (Valsiner, 2003; Cross, 2003; Stadler Elmer, 2015).
In short, songs are cultural rituals and devices that are used to share and to modulate affective states. Song singing also stimulates the learning of complex rules from early on, and hence, it seems to be the very first and most important cultural act for recognizing structures and rules, and for making abstractions.

IV. How can we describe and analyze vocal structures?

Song singing formally encompasses lyrics and melody or language and music. The two generative systems share the same organizational principle and vocalization as the expressive means. In order to describe and analyze song singing, it is important to summarize basic principles of the structures of children's songs. At the basis of Merker's work (2002) Stadler Elmer (2015) proposed a grammar of children's songs (CS) that makes explicit the underlying principles and rules. So far, the rules are restricted to German language since each language has specific prosodic rules that need to be taken into account (Hall, 2000). Up to now, seven general principles are formulated:

1) CS consists of lyrics (a verse or a poem) and a melody.
2) Two *generative systems* – music and language – are simultaneously involved that allow creating infinite possible songs.
3) The building block of both systems for making songs is the s*yllable* since it also contains musically relevant features such as pitch, time, and accents.
4) Songs are *hierarchical organized.*
5) Lyrics and melody are relatively autonomous (see 2), forming parallel hierarchies (cf. Baroni, Dalmonte, & Jacoboni, 1995). Ideally, lyrics and melody fit together in an equal manner. If this is not possible, one is super- and one is subordinated.
6) The poetic meter of the lyrics and the musical meter of the melody simultaneously rule the timing of the syllables. As aforementioned, this may create tension.
7) Symmetries of temporal and sonor forms aim at creating *well-formed wholes* or a Gestalt.

In order to understand the generative nature of the two systems, table 1 gives an overview on vocal features differently organized while singing, reciting, and speaking.

Table 1: Singing, reciting, and speaking share the same, yet differently organized vocal features. (© Stadler Elmer, 2015)

	Continuous dimension	Discrete categories (generative)	Common elements
singing	intensity	pitch, time	meter
reciting, chanting	intensity, pitch	time, phonems (vowels, consonants)	
speaking	intensity, pitch, time	phonems	phonems syllables

Note that poetic language – reciting or chanting – shares with singing the periodically accentuated pulse, a meter. Most notable is the fact that syllables are the commonly shared building blocks of the entire system, as formulated in principle 3. The traditional German song represented in figure 2 serves as an example to verify the given selection of rules. To date, 21 rules are formulated with respect to the tonal organization, the timing, and the lyrics (Stadler Elmer, 2015).

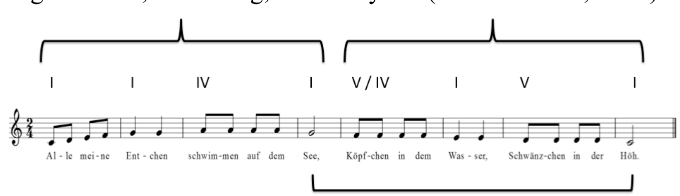


Figure 2: An example of a traditional German children's song. The top line shows the two phrases, the symbols I, IV, and V indicate the underlying harmonic structure of the melody, the last line points to the rhymes in the lyrics that mark the boundaries of the phrases. (© Stadler Elmer, 2015)

Timing rules - examples

1) Once the meter (measure) is set, it is valid for the entire song. CS don't change the measure or meter.
2) The number of measures in CS is dividable by two. Possible quantities of measures are 4, 8, 10, 12, 14, and 16.

Tonal rules – examples

1) CS are in *major keys*, rarely in pentatonic scale or minor keys.

2) The major key once chosen is maintained through the entire song. There is *no change of key*.
3) CS *begin* with one of the three tones of the tonic accord (do, mi, so).
4) The first measure of a CS marks the key.

Rules concerning lyrics - examples

1) The lyrics are formed in poetic language. That means it has a *meter*, and *verse lines* that end with *rhymes* (pair and cross rhymes).
2) The meter defines the periodic accents, and the verse lines is defined by the quantity of syllables.
3) The meter (trochee, jambus, dactylus, anapest) is contingent with the musical meter (measure) of the melody to yield a well-formed entity or Gestalt.

How can vocal learning be described and analyzed in its dynamic changes?

V. Micro-genetic analyses

In order to investigate children's strategies in generating new songs – be it invented or reproduced – we carried out a quasi-experimental study. We composed new songs that deliberately violated some of the CS rules, and we asked children to invent songs. Pictures stimulated all singing activities. In order to exemplify the application of CS rules, I use an excerpt of a previously published case study (Stadler Elmer, 2000, 2002) with Tom, 4;7 years old, and discuss it in the light of the CS grammar. Figure 3 shows Tom's first reproduction of a new song, he previously listened two times and joint in singing once. This graphic representation is based on a micro-genetic methodology including acoustic analysis by the Pitch Analyzer, and with the aid of the Notation Viewer (Stadler Elmer & Elmer, 2000).

Tom's first solo version (figure 3) shows first of all an exact match between his temporal organization and the model: Without counting 15, he produces 15 syllables that correspond to 15 notes or pitch categories, altogether forming a coherent and new song. He seems to have a sense for this formal aspect, or in other words, for the Gestalt of a song in terms of its temporal framework. He subordinates the lyrics by adopting only the first two words and by subsequently reducing the lyrics to the repetitive syllable */ne/*. The melody is remarkable: he adopts the first phrase, but he replaces entirely the second, unconventional one by a newly invented and well-formed tune. Thereby he generalizes the rule that a CS does not change the key, and therefore it ends with the tonic of the key introduced at the beginning. He continued acquiring this song by applying various strategies (adopting the lyrics, asking for help, focusing absolute pitch etc.).

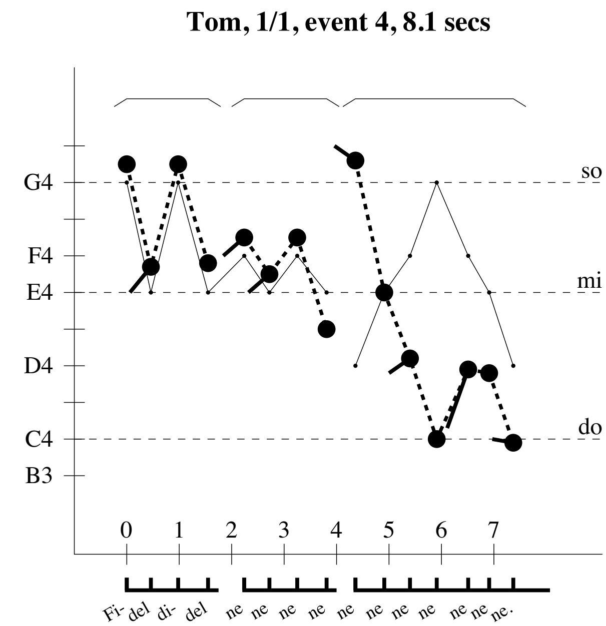


Figure 3: Tom's first version of a new song. The thin line represents the song model, and the dots represent pitches sung by the child. The *y*-axes shows pitch as a continuum, and the *x*-axes shows the continuous timing categorized as seconds, and measured by the onset of the syllable and duration. The horizontal lines indicating *do, mi, so* visualize the tonic triad as a conventional tonal framework (Stadler Elmer, 2000, 2002).

VI. Conclusions

The example given shows that song singing is more than a sequence of producing words, adding rhythm, then a melody contour, and finally intervals – one of the traditional conceptions of singing development (e.g., Welch, Sergeant, & White, 1998; Hargreaves, 1986; Moog, 1976). Microanalyses repeatedly reveal that children have an overall understanding for song singing being a coherently and densely organized social event with predictive features. It can't be acquired by adding elements or by focusing accurate pitch, but rather by grasping the entire form at the basis of related and repeated sensorimotor and simultaneously socially shared experiences.

References

Baroni, M., Dalmonte, R. & Jacoboni, C. (1995). The concept of hierarchy: A theoretical approach. In E. Tarasti (ed.), *Musical signification*. Essays in the semiotic theory and analysis of music (pp. 325–334). Berlin: Mouton de Gruyter.

Baumeister, R.F., & Leary, M.R. (1995). The need to belong: desire for interpersonal attachments as a fundamental human motivation. Psychol Bull 117(3):497–529.

Condon, W.S. (1977). A primary phase in the organization of infant responding. In: Schaffer HR (eds) Studies in mother-infant interaction. London: Academic, pp 153–176.

Cross, I. (2003). Music and biocultural evolution. In: M. Clayton, T. Herbert, R. Middleton (eds.) The cultural study of music: a critical introduction. London: Routledge, pp 19–30.

Dissanayake, E. (2000). Antecedents of the temporal arts in early mother-infant interaction. In: N.L. Wallin, B. Merker, & S. Brown (eds) *The origins of music* (pp 389–410). Cambridge, MA: MIT Press.

Dissanayake, E. (2008). The arts after Darwin: does art have an origin and adaptive function? In: K. Zijlmans, & W. van Damme, World art studies: exploring concepts and approaches, pp 241–263. Amsterdam: Valiz.
Dissanayake, E. (2011). Prelinguistic and preliterate substrates of poetic narrative. *Poetics Today 32*(1): 55–79. doi:10.1215/03335372–1188185
Eckerdal, P. & Merker, B. (2009). ‚Music' and the ‚action song' in infant development: An interpretation. In: S. Malloch, & C. Trevarthen (eds), *Communicative Musicality: Exploring the Basis of Human Companionship* (pp 241-262). Oxford: Oxford University Press.
Hall, T.A. (2000). *Phonologie: Eine Einführung.* Berlin: de Gruyter.
Hargreaves, D.J. (1986). *The developmental psychology of music.* Cambridge: Cambridge University Press.
Mampe, B., Friederici, A.D., Christophe, A., & Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Current Biology 19*(23): 1994–1997. doi:10.1016/j.cub.2009.09.064
Merker, B. (2009). Ritual foundations of human uniqueness. In: Malloch S Trevarthen C (eds) Communicative musicality: exploring the basis of human companionship. Oxford: Oxford University Press, pp 45–59.
Merker, B. (2002). Music: The missing Humboldt system. *Musicae Scientiae,* 6, 3-21.
Merker, B., Morley, I., Zuidema, W. (2015). Five fundamental constraints on theories of the origins of music. *Phil. Trans. R. Soc.* B 370: 20140095. http://dx.doi.org/10.1098/rstb.2014.0095
Moog, H. (1976). The development of musical experience in children of pre-school age. *Psychology of Music*, 4(2), 38–45.
Papoušek, H. & Papoušek, M. (1987). Intuitive parenting: a dialectic counterpart to the infant's integrative competence. In: Osofsky JD (eds) Handbook of infant development, 2 Aufl. New York: Wiley, pp 669–720.
Stadler Elmer, S. (2000). *Spiel und Nachahmung. Über die Entwicklung der elementaren musikalischen Aktivitäten*. Aarau: HBS Nepomuk.
Stadler Elmer, S. (2002). *Kinder singen Lieder: Über den Prozess der Kultivierung des vokalen Ausdrucks.* Münster: Waxmann.
Stadler Elmer, S. (2012a). Structural aspects of early song singing. In: Baldassare A (eds) *Music – Space – Chord – Image*. Festschrift for Dorothea Baumann's 65th birthday (pp 765–782). Bern: Peter Lang.
Stadler Elmer, S. (2012b). Infant vocal productions challenge music education. A case study on the transition between speaking and singing at age 14 months. *Proceedings of the Twenty-fourth International Seminar on Research in Music Education, Thessaloniki*, 223–231.
Stadler Elmer, S. (2015). Kind und Musik. Heidelberg: Springer.
Stadler Elmer, S. & Elmer, F.J. (2000). A new method for analyzing and representing singing. *Psychology of Music*, 28, 23–42. http://mmatools.sourceforge.net/mmatools.
Valsiner, J. (2003). Culture and human development: an introduction (Reprinted). London: Sage.
Welch, G.F., Sergeant, D.C., & White, P.J. (1998). The role of linguistic dominance in the acquisition of song. *Research Studies in Music Education*, 10, 67–73.

The Influence of Preliminary Listening and Complexity Level of Music on Sight-reading Performance

Yeoeun Lim,[*1] Seung-yeon Rhyu,[#2] Kyung Myun Lee,[+3] Kyogu Lee,[#4] Jonghwa Park,[*5] Suk Won Yi,[*6]

[*]*College of Music, Seoul National University, Republic of Korea*

[#]*Graduate School of Convergence Science and Technology, Seoul National University, Republic of Korea*

[+]*School of Humanities & Social Sciences, Korea Advanced Institute of Science and Technology, Republic of Korea*

[1]katieun@snu.ac.kr, [2]rsy1026@snu.ac.kr, [3]kmlee2@kaist.ac.kr, [4]kglee@snu.ac.kr, [5]koreadaniel@snu.ac.kr, [6]sukwon@snu.ac.kr

ABSTRACT

A major issue in sight-reading research is improving sight-reading proficiency. Previous studies have demonstrated the beneficial effect of repetitive aural rehearsal and input stimulus of the score on sight-reading performance, but it remains to be investigated whether a single preliminary listening with a different level of musical complexity can improve sight-reading proficiency. A total of 32 professional pianists participated in the experiment. Participants played six musical pieces lasting approximately 48 seconds each, composed specifically for this study. The musical pieces had three different complexities (low, medium, and high), and of the six pieces, participants listened to three—one each of low, medium, and high complexity—before engaging in sight-reading. They did not listen to the remaining three. Sight-reading performance was evaluated by calculating the distance between reference and performance data using dynamic time warping. A two-way ANOVA (with factors of listening and complexity) revealed a significant main effect of musical complexity ($p < .001$) on sight-reading performance, but no main effect of preliminary listening. Although preliminary listening had no effect on sight-reading proficiency, our results corroborate the importance of musical complexity on sight-reading. Further research should be conducted to investigate the influence of preliminary listening on different groups, such as novice sight-readers, in the same complexity condition.

I. INTRODUCTION

Sight-reading refers to playing a musical score with no rehearsal, and it is a basic and essential performance skill for every musician practicing and seeking to master music (McPherson, 1995). Many performers, including soloists and accompanists, are continually attempting to improve their sight-reading proficiency. Despite the plethora of empirical research on the process of sight-reading (Lehmann & Kopiez, 2009) and predictors of sight-reading success (Kopiez & Lee, 2006; 2008; Lehmann & Ericsson, 1996), practical methods for improving sight-reading proficiency have received relatively little attention. Several studies have demonstrated that listening to music before engaging in sight-reading influences sight-reading proficiency. Specifically, studies utilizing a cross-modality paradigm showed that preliminary listening before a sight-reading task helped musicians plan their motor execution (i.e., facilitated fingerings; Drai-Zerbib, Baccino, & Bigand, 2012). Additionally, passive listening to familiar music before sight-reading has been found to enhance motor performance without physical training in non-musicians (Lahav, Boulanger, Schlaug, & Saltzman, 2005).

Taken together, it can be speculated that auditory memory influences visual and motor processes, and that interactions between these modalities facilitate integration of information from scores during sight-reading. Two points, however, remain to be investigated: whether preliminary listening is effective when provided only once without repetition, and how the effect of preliminary listening varies depending on the input of sight-reading materials. Although there is clear evidence that complexity of music is a predictor of sight-reading skills (Kopiez & Lee, 2006; 2008) and sight-reading proficiency (Rayner & Pollatsek, 1997; Wurtz, Mueri, & Wiesendanger, 2009), such complexity has not been fully considered in a quantitative manner.

In this paper, we present a method of quantifying musical complexity, and determine how preliminary listening without repetition differently influences sight-reading performance according to the complexity of the music to be read. We hypothesize that listening to music before a sight-reading task with no repetition will have a positive effect on sight-reading proficiency, and that regardless of preliminary listening, the biggest difference in sight-reading proficiency will be present at the highest level of musical complexity.

II. METHOD

A. Participants

Thirty-two undergraduate students (28 female and 4 male; age range 18-24 years; mean age = 20.17 years) who majored in piano at Seoul National University participated in the experiment. Participants had begun their piano training at 5.93 years ($SD = 0.923$) of age, on average, and had played the piano for 15.11 years ($SD = 2.149$). They were thus regarded as having a high level of expertise in piano performance, and were familiar with sight-reading.

B. Materials

Six musical pieces of three varying levels of musical complexity (low, medium, and high) and two listening conditions (preliminary listening and non-listening) were composed specifically for this study. Each piece had the same key (C Major), time signature (4/4 meter), length (16 measures, 48 seconds), and tempo (80 BPM). The complexity of the music was determined according to difficulties in the rhythm and pitch, such that the overall complexity of the music was directly proportional to the levels of rhythm and pitch difficulties (represented as duration and accidentals, respectively): in other words, the shorter the duration and more accidentals that notes had, the higher the level of complexity of the piece.

To reduce the possibility of unaccounted-for dependent variables (i.e., other than preliminary listening and complexity of music), each piece was composed in a quantitative way (e.g., the total number of notes was matched within each complexity level; the number of accidentals and the longest duration of notes were more than doubling) (see Table 1).

Table 1. A quantitative composition of musical pieces: mm means measure numbers and indicates the number of notes on each measure; total notes, accidentals, and half notes mean the number of total notes, accidentals, and notes with half notes or over half notes, respectively.

	Complexity level											
mm.	**Low**				**Medium**				**High**			
1–4	10	8	8	10	8	7	11	8	8	10	12	10
5–8	8	9	9	8	9	10	12	8	13	12	10	12
9–12	6	10	7	7	11	10	11	11	14	13	18	13
13–16	8	6	9	7	12	11	11	10	11	10	12	12
Total notes	130				160				190			
Accidental	Null				27				84			
Half note	19				8				Null			
Key	C Major											
Meter	4/4											
Length	16 measures; 48 seconds											
BPM	80											

C.Apparatus

Musical pieces were written using the music notation software MuseScore 2.0.2, and sound files for the preliminary listening were created using Logic Pro X 10.2.2. Participants sight-read and played the scores on a YAMAHA CLP-525 Clavinova digital piano, and their performances were directly recorded in MIDI format.

D. Procedure

Out of the six musical pieces, participants listened to three—one each from the low, medium, and high complexity pieces—before engaging in sight-reading. They did not listen to the remaining three. The order of listening and playing was randomly and counter-balanced across participants. During the sight-reading performance, a metronome was given for every single beat to help participants keep a constant tempo. Participants listened to the metronome beats for two measures before performing each score, and were instructed to keep a steady tempo as much as possible.

III. RESULTS

A.Evaluation: Dynamic Time Warping

We utilized a computational method to examine how closely participants reproduced the sight-read materials. In other words, we calculated the overall distances between each of the performances and original pieces using dynamic time warping (DTW). DTW is a method to find an optimal cost path by tracking minimum costs within a similarity matrix of two sequences that vary in time and speed (Soulez, Rodet & Schwarz, 2003; Müller, 2007). It has been used effectively in the fields of speech or text recognition, as well as in score following. We applied this method to provide a single value reflecting the distance of the two different streams of music that should have been similar.

We divide the process DTW into two sections in this paper: computing the similarity matrix and calculating the DTW distance. First, we applied the Jaccard distance to obtain a similarity matrix between each participant's performance and the stimuli. The Jaccard distance is widely used for managing asymmetric binary variables (Willett & Barnard, 1998; Lipkus, 1999). The Jaccard distance between binary strings A and B can be calculated as following:

$$d_j(A,B) = 1 - J(A,B) = \frac{|A \cup B| - |A \cap B|}{|A \cup B|} \tag{1}$$

where $J(A,B)$ is the Jaccard similarity coefficient. To compute this distance between each performance and the reference, each MIDI file was initially converted into binary matrices comprising values of 0 and 1. The first column of the performance matrix and the first column of the reference matrix were then placed into inputs A and B of Equation (1), respectively. The remaining columns of each matrix were treated in the same manner, thereby producing a group of Jaccard similarity coefficients. The Jaccard similarity coefficients of each matrix formed a single similarity matrix.

Subsequently, the DTW distance was computed from each similarity matrix, the process of which can be represented as the DTW algorithm. This algorithm searches for the local minimum among the peripheral coordinates of each element of the matrix, including itself. When the local minimum is found, the value of the minimum is added to the element on the diagonal side. In other words, the values of the similarity matrix are filled with new values, which ultimately forms a new warping matrix. This process is depicted in Figure 1, and can be summarized in Equation (2):

$$\mathrm{C(i,j)} = \min \begin{cases} C(i,j-1) + D(i,j) \\ C(i-1,j) + D(i,j) \\ C(i-1,j-1) + D(i,j) \end{cases} \tag{2}$$

where $C(i,j)$ is a coordinate of the warping matrix, and is the coordinate of the previous similarity matrix. The location of each local minimum indicates the direction in which the cost path should proceed within the matrix, starting from the zero point. This path represents the optimal path for matching the two matrices. Particularly, the coordinates of the local minimum form the optimal cost path of the warping matrix, which is described in Figure 2. The value of the last coordinate of the optimal path is equal to the DTW distance. The distance for each performance was collected as raw data for the evaluation.

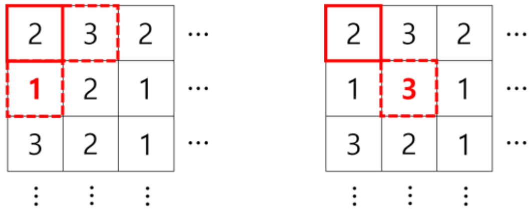


Figure 1. An example of the process for obtaining the warping matrix from the similarity matrix. The figure on the left indicates how to find the local minimum among a particular element (bolded border in the top-left corner) and its peripheral elements (dotted border) in the similarity matrix. In this figure, 1 is the local minimum. The figure on the right depicts how the local minimum is added to the element diagonal to the particular element (in the top-left corner).

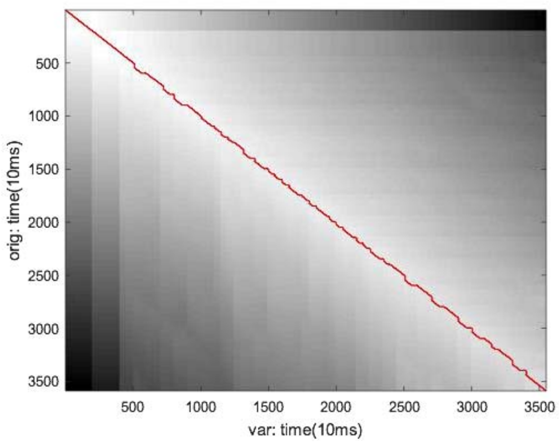


Figure 2. The warping matrix created from the DTW algorithm. The red line represents the optimal cost path of the warping matrix.

B. Results

We performed a two-way ANOVA on the data from the DTW distance to investigate the effects of preliminary listening and complexity of music on sight-reading. The results are described in Figure 3. Particularly, the bar plot in Figure 3 shows that the overall distance between the performance and original piece decreased with the level of complexity when participants previously listened to the music (Low level: 0.0818 to 0.0804; Medium level: 0.1126 to 0.1027; High level: 0.2192 to 0.2076; non-listening and preliminary listening conditions, respectively). However, the amount of decline within each complexity level was not significant ($F_{1,186} = 1.065, p > .05, MSE = 0.003$). Figure 3 also depicts how the overall distance increases with the complexity of music, regardless of listening condition (Low level: Mean = 0.0811; Medium level: Mean = 0.1076; High level: Mean = 0.2134). Furthermore, the differences between complexity conditions were significant ($F_{2,186} = 119.16, p < .001, MSE = 0.314$). This indicates that complexity of music significantly influenced sight-reading performance, regardless of whether players previously listened to music. The interaction between listening condition and complexity level was not significant ($F_{2,186} = 0.178, p > .05, MSE = 0.0005$).

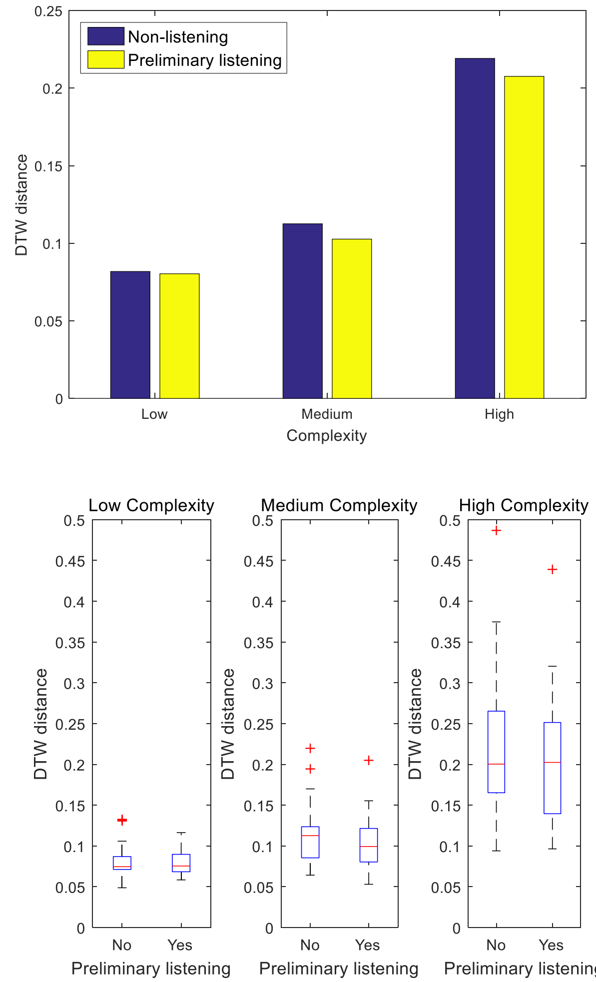


Figure 3. Bar plots (top) and boxplots (down) showing the effect of preliminary listening and complexity of music on sight-reading in the experiment. The DTW distance has no significant unit, but a relative value normalized between 0 and 1.

IV. CONCLUSION

We investigated the influence of preliminary listening and complexity level of music on sight-reading proficiency. In particular, the aim was to explore (1) whether a single preliminary listening can improve sight-reading proficiency and (2) how sight-reading performance differs according to the complexity level of music. Results revealed a significant main effect of music complexity on sight-reading proficiency. However, preliminary listening did not produce a significant effect.

Preliminary listening was investigated through performers anticipating the next movement of rhythm and pitch of music by auditory memory. The methodology was based on the assumption that in the non-repeated condition—listening to music only once without repetition before sight-reading—auditory memory does not aid visual and motor processes during sight-reading performance.

Sight-reading proficiency significantly differed depending on the complexity level of the music. Regardless of preliminary listening experience, the accuracy of sight-reading performance decreased as the level of musical complexity increased. Interestingly, the difference in sight-reading proficiency between the medium and high complexity levels was greater than that between the low and medium musical complexity levels.

We will explore the boundaries of auditory memory and aim to quantify the number of times auditory memory helps sight-reading proficiency in the future research. Further investigation of the influence of preliminary listening will be also conducted to examine the effect on different groups, such as novice sight-readers, in the same complexity condition. Examination of the effect of preliminary listening by measuring eye movements would further extend the body of knowledge in the field.

REFERENCES

Burman, D. D., & Booth, J. R. (2009). Music rehearsal increases the perceptual span for notation. *Music Perception: An Interdisciplinary Journal, 26(4),* 303-320.

Drai-Zerbib, V., Baccino, T., & Bigand, E. (2012). Sight-reading expertise: Cross-modality integration investigated using eye tracking. *Psychology of Music, 40(2),* 216-235.

Kopiez, R., & In Lee, J. (2006). Towards a dynamic model of skills involved in sight-reading music. *Music education research, 8(1).* 97-120.

Kopiez, R., & In Lee, J. (2008). Towards a general model of skills involved in sight-reading music. *Music education research, 10(1).* 41-62

Lahav, A., Boulanger, A., Schlaug, G., & Saltzman, E. (2005). The Power of Listening: Auditory-Motor Interactions in Musical Training. *Annals of the New York Academy of Sciences, 1060(1),* 189-194.

Lehmann, A. C., & Kopiez, R. Sight-reading. In S. Hallam, I. Cross & M. Thaut (Eds.) (2009). *The Oxford handbook of music psychology* (pp. 344-351). Oxford: Oxford University Press.

Lehmann, A. C., & Ericsson, K. A. (1996). Performance without preparation: structure and acquisition of expert sight-reading and accompanying performance. *Psychomusicology, 15(1/2),* 1-29.

Lipkus, A. H. (1999). A proof of the triangle inequality for the tanimoto distance, *Journal of Mathematical Chemistry*, 26, 263-265.

McPherson, G. E. (1995). Five aspects of musical performance and their correlates. *Bulletin of the Council for Research in Music Education, 127,* 115-121.

Müller, M. (2007). Dynamic time warping. *Information retrieval for music and motion*, 69-84.

Rayner, K., & Pollatsek, A. (1997). Eye movements, the eye-hand span, and the perceptual span during sight-reading of music. Current Directions in Psychological Science, 6(2), 49–53.

Sloboda, J. A. (1984). Experimental studies of music reading; A review. *Music Perception, 2(2),* 222-236.

Soulez, F., Rodet, X. & Schwarz, D. (2003). Improving polyphonic and poly-instrumental music to score alignment. *Proceedings of International Society on Music Information Retreival*, 6-11.

Waters, A. J., Townsend, E., & Underwood, G. (1998). Expertise in musical sight reading: A study of pianists. *British Journal of Psychology, 89,* 123-149.

Willett, P., Barnard, J. M. & Downs, G. M. (1998). Chemical similarity searching, *Journal of Chemical Information and Computer Sciences*, 38, 983-996.

Wurtz, P., Mueri, R. M., & Wiesendanger, M. (2009). Sight-reading of violinists: Eye movements anticipate the musical flow. Experimental Brain Research, 194(3), 445–450.

Perception of "Stolen Time" in Music: An Approach to the Systematization of Piano Performing

Rima Povilionienė
International Semiotics Institute, Kaunas University of Technology, Lithuania
rima.povilioniene@ktu.lt

ABSTRACT

Traditionally the concept of *rubato* marks the practice of playing with rhythmic freedom (speeding up / slowing down). Most often the irregularity of musical rhythm, meter, and / or tempo is not indicated in the score, but chosen by performers. Therefore the personally expressed features of *rubato* result in the uniqueness and variety of interpretations. This article presents an approach to the systemization of this indeterminate phenomenon in the practice of piano performing, taking into account the certain features of *rubato* as the indicators of a particular artist type. The different manifestations of playing freedom are analyzed in the case of Chopin's Nocturne No. 1 (B flat minor, Op. 9 No. 1), based on one computer generated sample-record and 42 different recordings (from 1928 to 2015, studio / live performances) by 32 pianists of different ages. The research focuses on the calculation and comparison of total time, different levels of disagreement in the vertical of the score, diversity of performing the groups of 11 and 22 eighth note-ornamentations, etc. The graphical representations of record extracts (mm. 1–4) were created using Sonic Visualiser 2.5. The examination of data obtained, audio and visual analysis of interpretations allows the determination of four particular groups of pianists, ranging from a quite stable and strict movement to a highly individual kind of narrativeness of sound.

INTRODUCTION

Traditionally the concept of *rubato* in music ("stolen time," derived from Italian *rubare* – "steal") marks the practice of performing with rhythmic freedom (speeding up and / or slowing down). A representative of Italian Baroque vocal art, Pier Francesco Tosi (~1653–1732) is regarded to be the first who used, in his treatise (1723), the term *rubare* (Hudson, 1997; Rosenblum, 1994), whereas discussing the ways of expressive singing: "Whoever does not know how to steal Time [*rubare il Tempo*] in Singing, knows not how to Compose, nor to Accompany himself, and is destitute of the best Taste and greatest Knowledge. | The stealing of Time, in the *Pathetick*, is an honourable Theft in one that sings better than others, provided he makes a Restitution with Ingenuity" (Tosi, 1743). However, the idea of stolen time as the liberty (or flexibility) of time in the style of singing was already emphasized a centenary earlier by another Italian author Giulio Caccini (1551–1618), who in *Le nuove musiche* (1601–02) mentioned *sprezzatura*, a kind of ornamentation, that enables the singers to express the emotion of the word without being restricted by the tempo: "Whence may appear that noble manner (as I call it) which, not submitting to strict time but often halving the value of the notes according to the ideas of the text, gives rise to that kind of singing with so-called "negligence" [*sprezzatura*]" (Caccini, 2009).

Usually in musical practice two conventional ways of expression of *rubato* are distinguished:

1) the flexible ornamentation of melody runs upon the regular accompaniment (e.g. in piano playing, the freedom of tempo and rhythm is achieved under the strict pulse of eighth-notes in arpeggios in the left hand in piano playing);
2) the flexibility manifests itself both in the melody and accompaniment (no accurate timing in both hands).

Both ways characterize the contrametrics in the vertical of musical time (contrametric *rubato*) rather than the flexibility of tempo using quickening and slowing (ritardando or accelerando, known as agogic *rubato*). Most often the irregularity of musical rhythm, meter, and / or tempo is not indicated in the score, but expressed by performers liberally (personally). Therefore the use of *rubato* results in the uniqueness of interpretation and a variety of artistic insights; as regarded by Polish pianist and pedagogue Paderewski, "Tempo Rubato is a potent factor in musical oratory, and every interpreter should be able to use it skillfully and judiciously, as it emphasizes the expression, introduces variety, infuses life into mechanical execution" (Finck, 1913).

BACKGROUND

This paper analyzes different manifestations of *rubato* in the case of Frédéric Chopin's (1810–1849) Nocturne in B flat minor, Op. 9 No. 1. The very first nocturne in Chopin's legacy of 21 examples of this genre was composed in 1830–1831 (Vienna and Paris), published in 1833, and dedicated to Maria Camillowa Pleyel, wife of the publisher and virtuoso pianist Camille Pleyel.

According to Mieczysław Tomaszewski, the composition "emerges from silence ..." (Tomaszewski, 2016). Actually the performance of this "emergency" in the first four bars may show a diversity of interpretations using the different approaches to the same musical text. In mm. 1–4 we face the irregularity of rhythm and subtly personified playing of ornamentations; the irregularity of motion is achieved in the right hand playing freely the various groups of eighth-notes (4, 6, 11, 22) whereas the left hand accompanies with eighth-note arpeggios.

The research was based on 42 recordings ranging from 1928 to 2015, among them – 29 studio and 13 live performances. The collected records were implemented by 32 pianists of different ages; some performers were examined from more than one record (see Table I). Different human records were analyzed taking into account the level of disagreement in the vertical of the score, calculation of total time and length of mm. 1–4 (because of highly varying tempo), performance of particular ornamentations etc. The human performances were compared to the computer generated sample-record based on the exact representation of original tempo (MM=116) and strict rhythmic pulse. The examination and comparison of record examples showed the gradual enrichment of interpretations with the features of *rubato* – manipulation with motion and

Table 1. List of Analyzed Records of Chopin's Nocturne B-flat Minor, Op. 9 No. 1.

No.	Pianist Name (in alphabetical order)	TT	mm. 1–4	Record Date	Live / Studio	Place / Label / Release, Re-mastering Date (if different from record date)
1.	Arrau Claudio (1903–1991)	5'41.499	18.662	1977-78	S	Amsterdam Concertgebouw / Philips Classics
2.	Ashkenazy Vladimir (b. 1937)	5'41.445	17.676	1997	S	Decca
3.	Barenboim Daniel (b. 1942)	5'38.644	15.544	1981	S	Studio Lankwitz Berlin / Polydor Int. GmbH / 1982
4.	Biret Idil (b. 1941)	5'29.859	15.883	1990-92	S	Naxos (*Grand Prix du Disque Chopin*, 1995)
5.	Chaimovich Vadim (b. 1978) (1)	6'40.067	17.974	2006	L	*No data*
6.	-"- (2)	6'34.845	17.573	2007	L	*No data*
7.	-"- (3)	7'34.894	20.865	2009	S	Sheva / 2010
8.	Cziffra Georges (1921–1994)	5'56.818	17.090	1980-86	S	EMI
9.	Dang Thai Son (b. 1958)	5'09.676	15.244	1987	S	Victor JVC
10.	Engerer Brigitte (b. 1952) (1)	5'37.917	17.741	1993	S	Harmonia mundi / 2014
11.	-"- (2)	5'43.339	17.413	2010 Jun 5	L	5th Piano Festival, Saint-Germain-au-Mont-d'Or
12.	-"- (3)	5'57.277	19.214	2010 Oct 9	L	Nuit de l'âme, Théâtre du Beauvaisis, Beauvais
13.	Falvay Sandor (1949)	4'56.153	16.090	1989 Jun	S	Budapest / Naxos
14.	Freire Nelson (b. 1944)	5'24.732	15.698	2009 Dec	S	Decca / 2010
15.	Gavrilov Andrei (b. 1955) (1)	5'22.581	20.810	2008 Aug 17	L	Valldemossa Chopin Festival, Majorca
16.	-"- (2)	*No data*	22.789	2013 May 17	S	Fazioli Hall, Rome
17.	-"- (3)	5'57.885	23.554	2013 Sep 2	L	Green Cross Int. 20th Anniv., VictoriaHall, Geneva
18.	Godowsky Leopold (1870–1938)	4'03.993	13.545	1928 Jun 23	S	English Columbia Recordings
19.	Harasiewicz Adam (b. 1932)	5'22.047	16.646	1961 Jun	S	The Netherlands / Philips Classics / 1994
20.	Joao Pires Maria (b. 1944)	5'27.787	16.765	1995-96	S	Deutsche Grammophon
21.	Kapell William (1922–1953)	5'23.701	15.164	1945 Feb 28	L	Carnegie Hall, New York / RCA
22.	Li Yundi (b.1982) (1)	5'18.124	14.620	2010 Mar 1	L	Chopin's 200th Anniv., Warsaw
23.	-"- (2)	5'28.234	15.165	2010	S	EMI
24.	-"- (3)	4'59.275	14.237	2011 Apr 23	L	Beijing National Centre
25.	-"- (4)	4'51.013	14.634	2015 Mar 1	L	Chengdu, China
26.	Magaloff Nikita (1912–1992)	6'41.030	20.204	1974	S	Amsterdam Concertgebouw / Philips Classics
27.	Moravec Ivan (1930–2015)	5'34.301	16.476	1965 Apr	S	St. Paul's Chapel at Columbia University, New York (Steinway) / Elektra Nonesuch / 1991
28.	Novaes Guiomar (1895–1979)	5'32.581	14.151	1957	S	Vox / 1995
29.	Ohlsson Garrick (b. 1948)	5'14.605	13.442	1979	S	EMI / 2005
30.	Pacini Sophie (b. 1991)	5'10.267	19.012	2014	S	CAvi
31.	Pollini Maurizio (b. 1942)	4'42.720	14.335	2005	S	Deutsche Grammophon
32.	Richter Sviatoslav (1915–1997) (1)	6'25.317	20.623	1950 Jun 19	L	Tchaikovsky Hall, Moscow
33.	-"- (2)	6'13.219	18.109	1972 Feb 16	L	Szeged, Hungary
34.	Rubinstein Arthur (1887–1982)	5'22.829	14.887	1965	S	RCA
35.	Sheng Cai (b. 1987)	5'29.595	15.254	2014 May 8	L	Capitol Theatre, New Brunswick, Canada
36.	Sombart Elizabeth (*No data*)	6'28.042	22.956	2012	S	Forlane / 2013
37.	Sztompka Fryderyk (1901–1964)	5'52.224	17.322	1959	S	Muza label (Polskie Nagrania)
38.	Thibaudet Jean-Yves (b. 1961)	5'16.165	16.975	2000	S	Decca
39.	Tiempo Sergio (b. 1972)	5'20.051	14.923	2006	S	EMI
40.	Tipo Maria (b. 1931)	6'50.769	21.560	1993-94	S	EMI
41.	Ts'ong Fou (b. 1934)	4'55.421	13.255	1978-80	S	Sony
42.	Varsi Dinorah (1939–2013)	4'35.471	13.661	1986	S	EMI

rhythmic freedom. This allowed the determination of four particular groups of pianists, ranging from quite strict and stable tempo motion to a highly individual kind of narrativeness of sound. Therefore the interpretations placed in the 1st and 4th groups are radically opposite, e.g. it is worth mentioning the records by the legendary pianist Leopold Godowsky (celebrated for the gorgeous singing tone, beautifully balanced voicing and masterful pedaling) versus Andrei Gavrilov (one of the most controversial, charming and unpredictable performers in the present piano world).

SYSTEMATIZATION OF PIANO PERFORMANCES

A computer generated sound-pattern was used as the starting point for analyzing the human records. The pattern is characterized by the original metrical implementation of MM=116, stable motion of eight-note arpeggios and proportional arrangement of irregular ornaments in the melody (mm. 2–3) with 11 and 22 eight-notes. It was supplemented with a graphical representation of melodic range spectrogram created by Sonic Visualiser 2.5. (Cannam, Landone & Sandler, 2010) (see Fig. 1). The estimated length of orderly performance of mm. 1–4 is 12.407” (the total duration of Nocturne performed by computer is 4’25.709”).

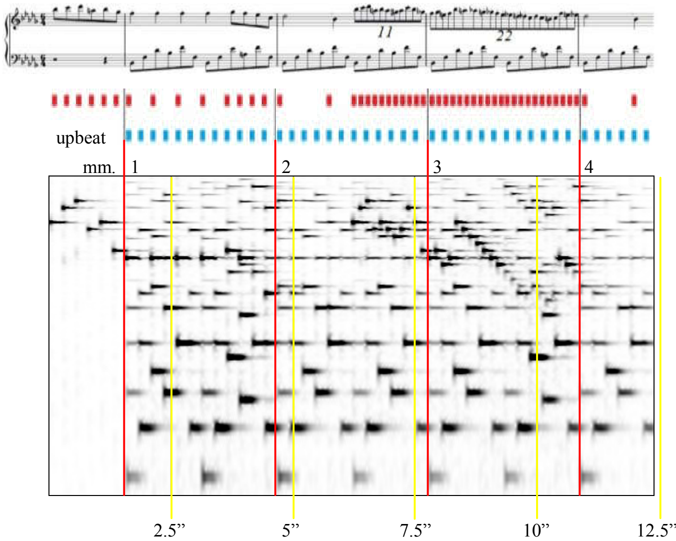


Figure 1. Visual representation of computer generated sound-pattern, MM=116, mm. 1–4 duration 12.407” (fragment of the score, scheme of quarter- and eighth-note sequences, and melodic range spectrogram)

On the first step of analysis, the first phrase in Chopin’s Nocturne (mm. 1–4, a traditional tonally closed period of two two-bar melodically and harmonically parallel phrases) was extracted from the total sound-record. Already here we encounter different approaches and the style of piano playing. Afterwards the extracts were grouped gradually moving from pretty strict playing to the very freedom of tempo and rhythm, which affected the manifestation of tempo from live to extremely slow motion. Therefore 4 types were distinguished, where No. 1 and No. 4 are absolutely opposite, with a high contrast of playing speed, which show a gradual change in slowing the tempo, and more and more mismatches in the metric vertical (see Table II).

The 1st type of performances are characterized by pretty orderly playing, stable tempo and equal motion of the accompaniment (eighth-notes in the left hand). The vertical of accompaniment and melody shows pretty strict metrical coincidences. In total 3 pianists/records were placed in this group: Falvay, Godowsky and Novaes. Usually the analyzed examples are united by live tempo, the length of performance of mm. 1–4 ranges from 13.545” to 16.090”.

The 2nd type: slight manifestation of *rubato* features. Enough live tempo and pretty strict pulse of the accompaniment; manipulation of minor acceleration and slowing down, especially inside the melodic ornaments of 11 and 22 notes; some records keep the strict pulse of the melodic phrase of 6 eighth-notes in the upbeat, whereas other tend to expose the upbeat (performed slightly slower and with more freedom, like the run-up); slight deviations in the match of the vertical; the second phrase (mm. 3–4) is performed more freely. 15 pianists/16 records: Rubinstein, Barenboim, Biret, Dang Thai Son, Freire, Kapell, Li Yundi (1, 2), Moravec, Ohlsson, Pollini, Sheng Cai, Sztompka, Ts’ong and Varsi. The length of mm. 1–4 is dominated by 14–15”. However, an equal motion, strict vertical and pretty stable rhythm is specific for extremely lengthy interpretation by Nikita Magaloff, lasting 20.204”.

Table 2. Duration Data of Sound-Records of Chopin’s Nocturne B-flat Minor, Op. 9 No. 1.

Length of mm. 1–4		
	MM=116	12.407”
1.	Ts’ong	13.255”
2.	Ohlsson	13.442”
3.	Godowsky	13.545”
4.	Varsi	13.661”
5.	Novaes	14.151”
6.	Li Yundi - 3	14.237”
7.	Pollini	14.335”
8.	Li Yundi - 1	14.620”
9.	Li Yundi - 4	14.634”
10.	Rubinstein	14.887”
11.	Tiempo	14.923”
12.	Kapell	15.164”
13.	Li Yundi - 2	15.165”
14.	Dang Thai Son	15.244”
15.	Sheng Cai	15.254”
16.	Barenboim	15.544”
17.	Freire	15.698”
18.	Biret	15.883”
19.	Falvay	16.090”
20.	Moravec	16.476”
21.	Harasiewicz	16.646”
22.	Joao Pires	16.765”
23.	Thibaudet	16.975”
24.	Cziffra	17.090”
25.	Sztompka	17.322”
26.	Engerer - 2	17.413”
27.	Chaimovich - 2	17.573”
28.	Ashkenazy	17.676”
29.	Engerer - 1	17.741”
30.	Chaimovich - 1	17.974”
31.	Richter - 2	18.109”
32.	Arrau	18.662”
33.	Pacini	19.012”
34.	Engerer - 3	19.214”
35.	Magaloff	20.204”
36.	Richter - 1	20.623”
37.	Gavrilov - 1	20.810”
38.	Chaimovich - 3	20.865”
39.	Tipo	21.560”
40.	Gavrilov - 2	22.789”
41.	Sombart	22.956”
42.	Gavrilov - 3	23.554”

Length of Total Piano Piece		
1.	Godowsky	4’03.993”
	MM=116	4’25.709”
2.	Varsi	4’35.471”
3.	Pollini	4’42.720”
4.	Li Yundi - 4	4’51.013”
5.	Ts’ong	4’55.421”
6.	Falvay	4’56.153”
7.	Li Yundi - 3	4’59.275”
8.	Dang Thai Son	5’09.676”
9.	Pacini	5’10.267”
10.	Ohlsson	5’14.605”
11.	Thibaudet	5’16.165”
12.	Li Yundi - 1	5’18.124”
13.	Tiempo	5’20.051”
14.	Harasiewicz	5’22.047”
15.	Gavrilov - 1	5’22.581”
16.	Rubinstein	5’22.829”
17.	Kapell	5’23.701”
18.	Freire	5’24.732”
19.	Joao Pires	5’27.787”
20.	Li Yundi - 2	5’28.234”
21.	Sheng Cai	5’29.595”
22.	Biret	5’29.859”
23.	Novaes	5’32.581”
24.	Moravec	5’34.301”
25.	Engerer - 1	5’37.917”
26.	Barenboim	5’38.644”
27.	Ashkenazy	5’41.445”
28.	Arrau	5’41.499”
29.	Engerer - 2	5’43.339”
30.	Sztompka	5’52.224”
31.	Cziffra	5’56.818”
32.	Engerer - 3	5’57.277”
33.	Gavrilov - 3	5’57.885”
34.	Richter - 2	6’13.219”
35.	Richter - 1	6’25.317”
36.	Sombart	6’28.042”
37.	Chaimovich - 2	6’34.845”
38.	Chaimovich - 1	6’40.067”
39.	Magaloff	6’41.030”
40.	Tipo	6’50.769”
41.	Chaimovich - 3	7’34.894”

The 3rd type: expression of “dreamy” *rubato*. A prominence of the accelerated or spaced out upbeat; a tendency to listen attentively to the melodic relief; rather slow tempo; acceleration and slowing down inside the pulse of eighths. 11 pianists/15 records: Arrau, Ashkenazy, Cziffra, Chaimovich (1, 2), Engerer (all 3 examples), Harasiewicz, Joao Pires, Li Yundi (3, 4), Richter (2), Thibaudet and Tiempo. The length of mm. 1–4 ranges from 14.237” to 18.662”.

The 4^{th} type: a turn towards the narrative nature of sound; maximally attentive listening and performance of each note that determines a very slow tempo; most often the vertical coincides in the beginning of bars only (the inside of bars is emphasized with the rhythmical irregularity); every performance is individual. 7 pianists/10 records: Chaimovich (3), Gavrilov (all 3 examples), Pacini, Richter (1), Sombart, and Tipo. The length of mm. 1–4 ranges from 17.573" to 23.554", dominated length over 20".

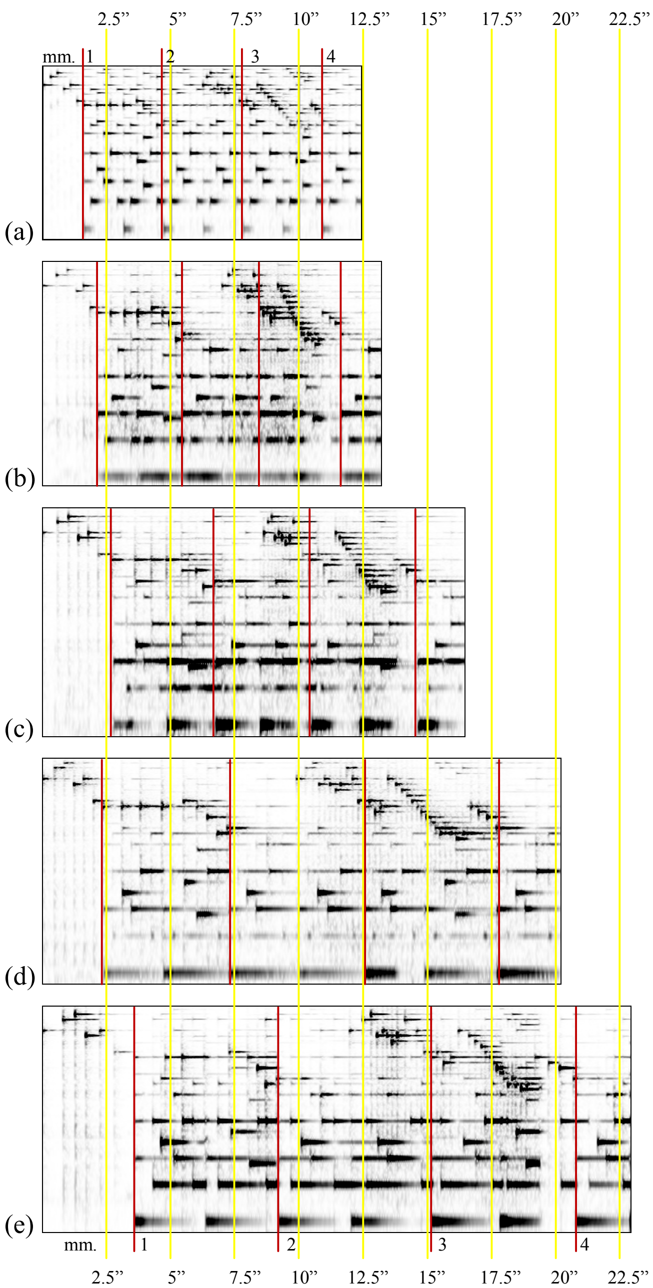


Figure 2. Some visual representations of Chopin's Nocturne in B-flat minor, Op. 9 No. 1, section mm. 1–4, by Sonic Visualiser 2.5.:

(a) – computer generated sound-sample 12.407"
(b) – Ts'ong 13.255"
(c) – Moravec 16.476"
(d) – Magaloff 20.204"
(e) – Sombart 25.956"

CONCLUSION

The main observation after examining the records highlights the manifestation of *rubato* closely related to the moderation of movement: the tendency towards the expressiveness of sound, an attempt to highlight every eight-note (likely sing) slows down the tempo. On the other hand, tempo deceleration is directly connected to the disagreement in the vertical between melody and accompaniment (see Fig. 2).

As expected, the interpretations, placed in the groups 2 and 3, constitute the major part of the records. The length of interpretation and choice of narrative approach (specific for group 4) does not depend on the age of performer or the date of the record. However, the most ordinary performances in the 1^{st} group are represented especially by pianists of the old generation (Leopold Godowsky and Guiomar Novaes). In most cases, the total duration of Nocturne is possible to predict according to the way of interpretation in mm. 1–4 (Pacini's performance displays unpredictable proportion: the lengthy period and narrative playing in mm. 1–4 lasts over 19", however the total duration of the piano piece is of an ordinary length 5'20").

In the range of 42 interpretations only a few pianists are close to the original tempo and duration, e.g., Ts'ong, Ohlsson, Godowsky, Varsi (see Table II). After comparing Varsi's interpretation to the Alfred Cortot's notes in his 1945-edition of Nocturne Op. 9 No. 1, indicating the tempo mark 116 and total duration of Nocturne 4'30", the Uruguayan pianist Dinorah Varsi may be regarded as the closest to the original conception of Nocturne tempo.[1]

ACKNOWLEDGMENT

The preparation of the article and participation at ICMPC14 partly supported by the Lithuanian Composers' Union and Kaunas University of Technology.

REFERENCES

Caccini, G. (2009). *Le nuove musiche.* H. Wiley Hitchcock (ed.), 2nd ed. Middleton, Wisconsin: A-R Editions.

Cannam, C., Landone, C., & Sandler, M. (2010). Sonic Visualiser: An Open Source Application for Viewing, Analysing, and Annotating Music Audio Files", in *Proceedings of the 18th ACM International Conference on Multimedia*, 1467–1468. Retrieved April 18, 2016, from http://sonicvisualiser.org/sv2010.pdf.

Finck, H. T. (1913). *Success in Music and How it is Won*. New York: C. Scribner's sons, chapter "Paderewski on Tempo Rubato".

Hudson, R. (1997). *Stolen Time: The History of Tempo Rubato*. Oxford: Oxford University Press.

Rosenblum, S. P. (1994). The Uses of Rubato in Music, Eighteenth to Twentieth Centuries. In *Performance Practice Review*, 7(1), article 3, 33–53.

Tomaszewski, M. (2016, April 18). *Series of Programs "Fryderyka Chopina Dzieła Wszystkie"*, Polish Radio, Program II. Retrieved April 18, 2016, from http://en.chopin.nifc.pl/chopin/composition/detail/id/153.

Tosi, P. F. (1743). *Observations on the Florid Song, or Sentiments on the Ancient and Modern Singers.* London: J. Wilcox.

[1] It is regarded, that Alfred Cortot (1877–1962) was highly influential for his poetic interpretations of piano pieces by Chopin and Schumann. As is known, Cortot studied at Paris Conservatoire with Chopin's student Emile Decombes. Respectively, the Uruguayan pianist Dinorah Varsi (1939–2013) studied under Sarah Bourdillon, a pupil at the Alfred Cortot's Ecole Normale de Musique. So, Varsi's interpretation of Chopin's Nocturne is a possible legacy of Cortot's and respectively Chopin's conception.

Micro manipulations, macro *affect*: is expressive accent perception influenced by auditory-biography in live music performance?

Cynthia-Louise Dellit[1]
[1]School of Creative Arts, University of Newcastle, Newcastle, New South Wales, Australia

Mystification continues to surround artistic personae and interpretive powers of performing musicians. Recent works suggest that attempting to quantify building blocks of artistic communication may evoke resistance from sections of the performance/teaching community although a more explicit approach could clearly benefit students. My novel research contributes to "demystification" of artistry by analysing/trialling in the area of non-notated accenting. Accent, described as a complex phenomenon, occupies an augmented expressive space involving micro-manipulations of multiple sound parameters and contributes significantly to the communication of musical *affect*. I have designated auditory-biography (AB), i.e. personal listening history, as both socially situated and shaped by specialized instrument/ensemble exposure. This paper reports on a study investigating influences of AB on perceptual accenting accuracy. Inter-musician perception of expressive accenting patterns was examined using stimuli from live Bach sonata performances. Listener participants were performance majors from diverse instrumental backgrounds. The three solo performers (contrasting instruments) were experienced professionals. Listeners marked on a score beat onsets which were highlighted in some way. The trial, led by the performer-as-researcher, explored the hypothesis: "A group of musicians will perceive the *same* sound cue as *different* sound cues with the difference indicating AB instrument/ensemble/social influences". Data was pre-processed using SDT, analysed with general linear models using inference via Bayes factors. Results confirmed extensive correlations between AB and accenting accuracy, including influences of timbral/instrument register familiarity, ensemble genre and socially situated bias. This study increases understanding of ensemble training programs as well as the multi-layered, richly interconnected areas of AB and accent as enacted during the interpretation of music texts.

The Seattle Singing Accuracy Protocol: An Automated Measure of Singing Accuracy

Steven Demorest[1], Peter Pfordresher[2]
[1]Bienen School of Music, Northwestern University, Evanston, IL, USA
[2]Department of Psychology, University at Buffalo, the State University of New York, Buffalo, NY, USA

The Seattle Singing Accuracy Protocol (SSAP) is a set of standardized tasks designed to measure singing accuracy and component skills including matching novel pitches and patterns in different timbres, singing a familiar song from memory on text and neutral syllable, and testing perceptual acuity. The SSAP was developed by a group of researchers from cognitive psychology, neuroscience, and education in an attempt to standardize the way certain basic tasks were being measured. The goal was to provide a measure of basic singing tasks that were short enough to be included in any study of singing accuracy thus allowing researchers to compare findings across studies and develop and database of singing performance across populations of different ages and skill levels. Our goal is to make the measure available to any scholars interested in studying singing accuracy and its development.

This presentation will provide a brief demonstration of the measure and focus on the methodological challenges inherent in creating an accessible version of the protocol for use online and the quality of the initial data we are receiving. Methodological challenges include insuring clarity of the tasks, quality of input, accurate real-time acoustic signal analysis, and a user-friendly interface. We will share an analytical comparison of how well the automated acoustic analysis mimics human judgments of accuracy for both pattern matching and a cappella singing. We will also offer a descriptive analysis of preliminary data collected from the general public since the beta version went live in November of 2015 to see who is choosing to participate and what kind of picture we are getting of singing accuracy in the general population.

A classical cello-piano duo's shared understanding with their audience and with each other

Neta Spiro[1,2] and Michael F. Schober[3]
[1]Research Team, Nordoff Robbins, London, UK
[2]Centre for Music and Science, University of Cambridge, Cambridge, UK
[3]Department of Psychology, New School for Social Research, The New School, New York, NY, USA

To what extent do a pianist and cellist understand what just happened in their live duo performance in the same way as each other, and in the same way as their audience members do? In previous case studies on jazz standard improvisation and free jazz improvisation, performers did not fully agree with their partner's characterizations of what occurred in the improvisations, and they could agree with an expert outsider's characterizations more than they agreed with their partner. Can this pattern also be observed in classical chamber performance? In the current study, a cello-piano duo performed Schumann’s Fantasiestücke Op. 73 no. 1 once for the other 12 members of their conservatory studio class, which included another duo who were prepared to play the same piece that day. Immediately afterwards, the listeners and players individually wrote the top three things that had struck them about the performance. Then, listening to the recording on personal devices, they marked on scores and wrote about three moments that struck them as most worthy of comment. After all had finished, the response sheets were redistributed for each participant to rate their agreement (on a 5-point scale) with, and comment on, another participant's characterizations. Following this, there was a second redistribution for agreement ratings with one other participant’s characterizations. Results showed that the performers did not select any of the same moments in the piece as worthy of comment and that they agreed with less than half of their performing partners’ statements; this was markedly less than they agreed with statements by the listeners who were prepared to play and markedly less than the other listeners’ agreement with each others’ statements. Even some agreements included comments demonstrating that the agreement was qualified. As we have seen in other genres, performers’ understanding of what happened is not necessarily privileged relative to outsiders’ understanding.

The Role of String Register in Affective Performance Choices

Caitlyn Trevor,[*1] David Huron[*2]

[*]*Music, Ohio State University, Columbus, USA*
[1]trevor.5@osu.edu, [2]huron.1@osu.edu

ABSTRACT

When aiming to increase emotional intensity, advanced string performers typically prefer playing a passage using a high position on a low string rather than playing the same pitches using a low position on a high string. Research in emotional speech offers a possible explanation for this practice. For both strongly negative and positive emotions, people tend to speak higher in their vocal range. Accordingly, utterances at the identical pitch level will sound more emotional if the speaker is perceived to have a lower tessitura than a higher tessitura voice. We theorized that the practice of moving to a high position on a lower string might mimic this vocal emotional communication practice as a way of conveying higher emotional intensity. Presented here are two studies to test the conjectured effect. The first study aims to determine whether listeners can reliably distinguish between tones played in high and low string positions. Twelve sustained cello tones were recorded in both first position on a higher string and fifth position on a lower string. Participants were briefed on the playing positions on a cello and given sound examples. Listeners were then asked to identify whether the recorded tones were played in first or fifth position. In the second study, we test the hypothesis that melodies played in fifth position are perceived as more emotionally intense than the same melodies played in first position. Six short melodies were recorded in both first and fifth positions. In a 2AFC paradigm, listeners chose which of the two versions of the melodies they perceived as more emotionally expressive. The results of this study demonstrate that listeners are reasonably able to differentiate between pitches played in low and high playing positions and might interpret melodies played in a high playing position as more expressive than those played in a low playing position.

I. INTRODUCTION

For string players, one of the important musical decisions that must be made when approaching a piece of music is the fingering, or how the string player will position their hand in order to play the notes on the score. There are two important aspects to consider when deciding on the fingering of a musical passage. One is playability. A certain location on the fingerboard, or hand position, could facilitate fingering a certain musical passage better than another, especially for a passage with many string crossings or chords. The second aspect to consider is musical expression. When interpreting musical passages, string players need to decide on an emotional intention for each musical moment, melodic line, section, etc. With an emotional intention in mind, a string player must then decide on how best to communicate it using expressive musical tools such as timbre, vibrato, and dynamics.

Relatedly, the observation was made that string players, specifically cellists, move the fingering up the fingerboard to a high playing position (away from the scroll towards the bridge) when they aim to make a passage sound more expressive. For this study, an ecologically based theory is presented that could explain this observed behavior.

When communicating more intense emotions (such as panic, elation, and despair), humans tend to speak higher in their range than for less intense emotions (such as contempt, boredom, and shame) (Banse & Scherer, 1996). Humans are highly effective at predicting where in one's voice someone is speaking (Honorof & Whalen, 2005). It may be possible to perceive this information from instruments as well. If one can hear cues of location in range from a string instrument, this could support why using high playing positions might communicate a higher level of emotional intensity. In this study, part one tests whether or not it is possible to recognize pitch register from an instrument. Part two then studies whether there is an effect for emotional intensity based on high or low playing positions.

For part one, the ability to hear pitch register on the cello was investigated by testing whether participants were able to categorize cello tones into high and low playing positions. Specifically, the following was hypothesized: *(H1) When asked to compare individual cello notes to recorded examples of high and low playing position notes, participants will be able to identify which notes are played in which registers*.

For part two, to test whether there is an emotional effect of register, participants were asked to listen to pairs of cello melodies (one played in a high playing position and one played in a low playing position) and decide which one was more expressive in a two-alternative forced-choice design. For part two, the following was hypothesized: *(H2) When exposed to a cello melody played once in the lower string register and once in the upper string register, participants will select the upper register version as the more expressive recording.*

II. METHOD

A. Participants

The 47 participants for this study were mostly undergraduate and graduate music majors from Ohio State University and Illinois Wesleyan University between the ages of 18-30. Additional participants were between the ages of 18-40 and were a mixture of musicians and non-musicians.

B. Materials

To ensure that the predicted phenomenon did not apply only to specific instruments or cellists, four cellists were recorded playing the notes for part one and two cellists for the melodies for part two. In string performance, another factor that influences timbre is whether a given pitch coincides with a harmonic of an open string. The resulting resonance due to sympathetic vibration can produce a marked change in timbre. Accordingly, both sympathetic pitches and non-sympathetic pitches were used as stimuli. Specifically, half of the stimuli involved sympathetic pitches and half were non-sympathetic pitches. The selected pitches also covered a large range on the cello and used each of the four strings. Each pitch had a low and high playing position note. Twelve pitches total were selected for part one of this study. Six melodies were selected from *Suzuki Cello School: Cello Part,* Volumes 3 and 4 (1999, 2003). It was deemed practical to equally represent both major and minor mode because of their opposite emotional connotations. Therefore, three of the pieces picked were in a major mode and three were in a minor mode. The minor mode is often associated with sadness (Post & Huron, 2009). It was observed that cellists might move into the upper register more often for a sad expressive musical passage than for a happier one. Therefore, the prediction was made that participants would select the high playing position melodies as more expressive more often for the minor mode selections than major mode selections: *(H3) Participants will select the melodies played in a high playing position as more expressive for minor mode melodies more often than for major mode melodies.*

C. Interface

JavaScript and Jspsych were used to create the interface for this study (Flanagan, 2006; De Leeuw, 2015). When participants clicked on the first link for part one, a main window and a pop-up window appeared. The pop-up window contained the audio examples of notes played in a high playing position and notes played in a low playing position. The main window thanked them for participating and then proceeded to the demographic questions. After these, participants were briefly informed of what playing positions were on the cello and asked to familiarize themselves with the examples in the pop-up window as much as they liked throughout that part of the study. They were then presented with a pair of notes and asked to identify which they believe was played in a *high playing position* by pressing 1 or 2 on their computer keyboard. For part two, a similar procedure occurred. Participants were presented with pairs of melodies and asked to select which performance they deemed more expressive.

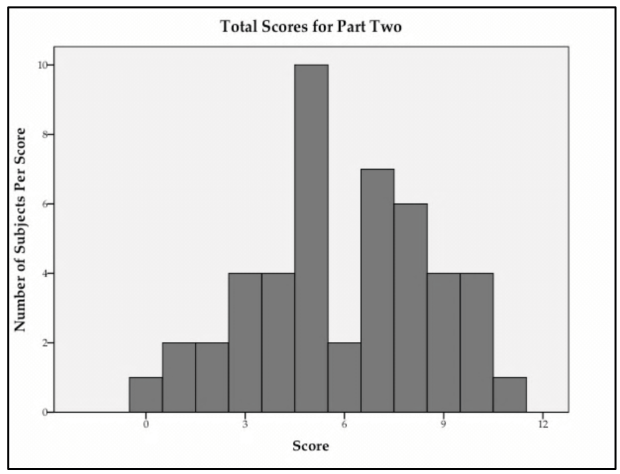
III. RESULTS

Recall that the first hypothesis predicted that participants would be able to hear a difference between the same notes played in a low playing position versus a high playing position. This ability was investigated by asking participants to listen to a pair of notes and choose which one was played in a high playing position. Each participant heard 12 pairs of notes. Therefore, if the results were consistent with the hypothesis, the resulting mean of scores should be higher than chance (6). A simple *t*-test demonstrates that the results are consistent with our hypothesis in that participants on average scored significantly higher than 6 ($M = 7.96$, $SD = 2.86$, $t\,(46) = 4.69$, $p < .0001$).

Our second hypothesis predicted that melodies played in mostly upper playing positions would be perceived as more expressive than those played in lower playing positions. To test this, participants heard 12 pairs of melodies, one version played with higher playing positions and one with lower playing positions. They were asked to select which of the two was more expressive. The results were coded so that selecting the melody played in an upper playing position was a "correct" answer. Therefore, should the results support the hypothesis, the mean score should be greater than 6. Initially it could appear that our results are inconsistent with our hypothesis (M=5.96, SD=2.7). However, upon closer examination, these results still provide an interesting picture.

Figure 1 shows a distribution of the frequencies of the results for part two of this study. As you can see, there are two swells in the data. Participants either scored mostly below 6 (showing a preference for melodies played in a low playing position) or mostly above 6 (showing a preference for melodies played in a high playing position). Few participants actually scored around chance (6). To test the significance of this phenomenon, a *post-hoc* test was run. To verify that the means of each half of the data were significantly different

Figure 1. The Distribution of Scores for Part Two



than chance, six was subtracted from each of the scores and the absolute value was taken. A *t*-test was used again, this time with a test value of zero since zero represented participants lacking a strong preference for either low or high playing positions. The results are consistent with the idea that participants had an opinion that either low playing position melodies or high playing position melodies were more expressive rather than playing position having no affect ($M = 2.28$, $SD = 1.44$, $t\,(46) = 10.84$, $p < .0001$).

Recall that for this part of the experiment, two cellists were used. Half of the participants heard one cellist (Cellist A) play all of the melodies and half of them heard the other (Cellist B). Comparing resulting scores with an independent means *t*-test shows that there is a significant difference between responses to melodies played by Cellist A ($M = 7.67$, $SD = 1.99$) and responses to melodies played by Cellist B ($M = 4.17$, $SD = 2.15$, $t\,(45) = 5.78$, $p < .0001$). This difference is also evident in the distributions presented below further supporting the results from the *post-hoc* test (Figures 2 & 3). Participants who heard Cellist A showed a significant preference for melodies played in the upper register ($t\,(23)=4.1$, $p < .0001$) and participants who heard Cellist B showed a significant preference for melodies played in the lower register ($t\,(22) = -4.1$, $p < .0001$). Theories behind this difference in results are examined in the discussion section.

The third hypothesis predicted that participants would select melodies played in a high playing position as the more

Figure 2. The Distribution of Scores for Cellist A

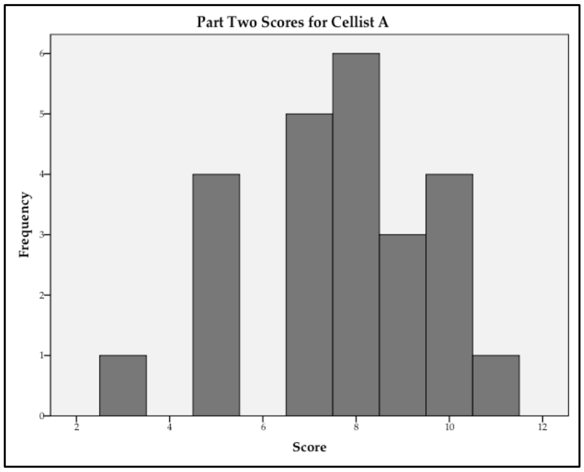
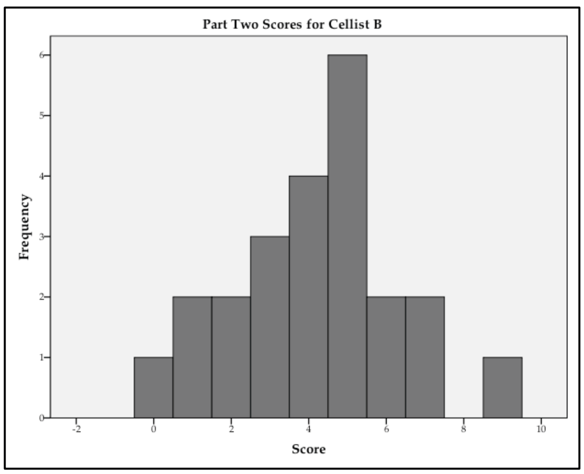


Figure 3. The Distribution of Scores for Cellist B



expressive melody if the melody was in the minor mode rather than in the major mode. An independent means *t*-test was conducted. The results were consistent with the hypothesis in that, for both cellists, melodies in the upper playing positions were rated as more expressive more often when they were in the minor mode (M = 3.4, SD = 1.66) than in the major mode (M = 2.53, SD = 1.6, t (92) = -2.59, p = .011).

IV. DISCUSSION

The results of the first part of the study were consistent with the prediction that listeners would be reasonably able to categorize cello pitches into high and low playing positions. In additional analyses of the data for part one, no effect was found for what instrument the listener played or for whether or not the tones played were sympathetic or non-sympathetic pitches. The results of the second part of the study were partially consistent with the second main hypothesis. Participants who heard recordings by Cellist A, on average, rated the melodies played in high playing positions as more expressive. However, participants who heard Cellist B's recordings, on average, rated the melodies played in low playing positions as more expressive. Given that the ability to play musically successfully in the upper register of the cello requires a high level of skill, it is worth noting that Cellist A was at a more advanced playing level than Cellist B. Cellist A had just completed a bachelor's degree in cello performance and was applying to graduate programs. Cellist B had just finished their sophomore year of undergraduate studies in cello performance. This could perhaps explain the difference in results. Looking at the written responses, both the highest scores for each cellist indicate that dynamic variation, timbre, rubato, and vibrato were all factors in their responses. This might indicate that Cellist A was able to execute these variables well in a high playing position while Cellist B was less able to. For Cellist B, responses also mentioned intonation and emotion as part of their response possibly indicating that in the upper register, Cellist B played less in tune and more robotically. Additionally, perhaps results would have been different had the cellists been able to practice the melodies before the recording session.

Considering the positive results for Cellist A, it is worthwhile to discuss other reasons why cellists might prefer the upper register of the instrument when aiming to play more expressively. One reason is that vibrato changes as one moves up the fingerboard. In some empirical research, vibrato has been found to increase in both width and rate when moving from a low playing position to a high one. Allen, Geringer, & MacLeod (2009) found that the rate of vibrato of a professional violinist increased from a mean of 5.7 Hz in first position to a mean of 6.3 Hz in fifth position, and that their vibrato width increased from 40 cents in first position to 108 cents in fifth position. This finding extends beyond that single-case example. MacLeod (2008) found in another study with high school and university violin and viola players that the vibrato width was wider in the upper register with a mean of 58 cents as compared to 34 cents in the lower register, and that the rate was also faster in the upper register. Similarly, David A. Pope (2012) measured vibrato characteristics of high school and college level cellists and found that vibrato rates increased from a mean of 5.07 Hz in first position to a mean of 5.33 Hz in fourth and thumb positions, and that the width increased from 23 cents in first position, to 34 cents in fourth position, to 43 cents in thumb position.

The explanation for why vibrato widths and rates increase as one's hand moves up the fingerboard is partially related to physics while also being effected by cello technique and characteristics about the instrument. Generally, vibrato width increases as one moves up the string because the string length shrinks causing the physical distance between the notes to diminish as well (MacLeod, 2008). Because of this, pedagogues generally encourage students to use a smaller vibrato width as they move up the instrument. Executing a thinner vibrato width results in a reduction of the width of the oscillating motion of the hand. Consequently, the time it takes to complete a full oscillation is also reduced resulting in a faster vibrato rate. This phenomenon is further complicated by changes that hand position and vibrato technique go through as one moves to higher playing positions. For example, when moving into fourth position, the neck of the instrument ends and the body of the instrument partially obstructs the hand as it vibrates thereby shrinking the vibrato width and rate. The consequential vibrato alterations of shifting hand positions may factor into why a string player would opt for one hand position over the other for expressive reasons.

In order to predict why a string player might chose a wider, faster vibrato, it is necessary to consider what the emotional impressions are of various vibrato widths and rates. Research on vocal vibrato, which is comparable in range and width to string vibrato (Seashore, 1931), gives insight into the affective information of vibrato widths and rates. Dromey, Holmes, Hopkin, & Tanner (2015) analyzed recordings of graduate vocalists and found increased vibrato rates for selections that were rated as more emotionally intense. Other studies by Sundberg (1994) & Sundberg *et al.* (1995) found the same correlation between vibrato rate and emotional intensity as well as some suggestion of vibrato width being positively correlated as well. Assuming that vibrato could carry the same emotional meaning for instruments other than voice, perhaps string players move to higher playing positions for a wider, faster vibrato to give a passage a higher emotional intensity.

Given this research, it is hard to say which would be the dominant reason for moving into the higher register: to use a high pitch register to portray high emotionality or to achieve a wider, faster vibrato to portray high emotionality. Also, it may be that the two are intertwined. Perhaps a wider, faster vibrato is part of the cue for high register, which communicates high emotionality. On the other hand, perhaps high register portrays a greater emotionality because of the wider, faster vibrato. This study cannot provide a conclusive answer to this relationship. Perhaps future work will be able to differentiate between these possible factors and determine any possible causality.

Another potentially expressive device, timbre, may be affected by fingering decisions as well. Open (un-fingered) strings have been found to have a brighter timbre than stopped (fingered) strings because the finger on the string dampens the higher harmonics (Schelleng, 1973). If a string player wants to avoid the brightness of the open strings, they may be tempted to move to a higher playing position to facilitate playing the closed versions of those notes. There may also be a timbral difference between notes played on a high string in a low playing position and those played on a low string in a high playing position. No measurements of timbre-related acoustical features were taken for this paper. However, in the future, it would be helpful to examine the spectral differences of each register given that our results indicate an audible difference.

Results were also consistent with our third hypothesis, that participants would more often prefer upper playing position melodies when in the minor mode compared to the major mode. It could be that playing in the upper register has a darker timbre than in the low register. A darker timbre has been found to be an element of sad speech (Scherer, Johnstone, & Klasmeyer, 2003). The minor mode is also associated with a sad affect (Juslin & Sloboda, 2011). Therefore, perhaps if the upper register indeed has a darker timbre, participants found the sadder minor mode melodies to be more suited to that upper playing position.

As mentioned earlier, it is impossible to declare exactly which musical element of the upper range of the cello might have contributed to the perceived higher level of emotional intensity for Cellist A. While vibrato and timbre are certainly a part of the equation, both parts of this study provide strong evidence for the conjecture that using high playing positions on the cello might communicate a higher emotionality by mimicking the similar vocal affective cue.

ACKNOWLEDGMENT

Our thanks to the members of the Cognitive and Systematic Musicology Laboratory at Ohio State University for their critical feedback and support.

REFERENCES

Allen, M. L., Geringer, J. M., & MacLeod, R. B. (2009). Performance practice of violin vibrato: An artist-level case study. *Journal of String Research*, *4*, 27-38.

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of personality and social psychology*, *70*(3), 614.

Bharucha, J. J. (1991). Pitch, harmony, and neural nets: A psychological perspective. In P. Todd & G. Loy (Eds.), *Music and Connectionism* (pp. 84-99). Cambridge, MA: MIT Press.

De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1-12.

Dromey, C., Holmes, S. O., Hopkin, J. A., & Tanner, K. (2015). The Effects of Emotional Expression on Vibrato. *Journal of Voice*, *29*(2), 170-181.

Flanagan, D. (2006). *JavaScript: the definitive guide*. " O'Reilly Media, Inc."

Honorof, D. N., & Whalen, D. H. (2005). Perception of pitch location within a speaker's F0 range. *The Journal of the Acoustical Society of America*, *117*(4), 2193-2200.

Juslin, P. N., & Sloboda, J. (Eds.). (2011). *Handbook of music and emotion: Theory, research, applications*. OUP Oxford.

MacLeod, R. B. (2008). Influences of dynamic level and pitch register on the vibrato rates and widths of violin and viola players. *Journal of Research in Music Education*, *56*(1), 43-54.

Pope, D. A. (2012). An Analysis of High School and University Cellists' Vibrato Rates, Widths, and Pitches. *String Research Journal*, *3*, 51.

Post, O., & Huron, D. (2009). Music in minor modes is slower (except in the Romantic Period). *Empirical Musicology Review*, Vol. 4, No. 1, pp. 1-9.

Seashore, C. E. (1931). The natural history of the vibrato. *Proceedings of the National Academy of Sciences*, *17*(12), 623-626.

Schelleng, J. C. (1973). The bowed string and the player. *The Journal of the Acoustical Society of America*, *53*(1), 26-41.

Scherer, K.R., Johnstone, T., & Klasmeyer, G. (2003). Vocal expression of emotion. In: R. Davidson, K. Scherer, & H. Goldsmith (eds). *Handbook of Affective Sciences*. Oxford: Oxford University Press, pp. 433-456.

Sundberg, J. (1994). Acoustic and psychoacoustic aspects of vocal vibrato. *STL-QPSR, 35*(2-3), 45-68.

Sundberg, J., Iwarsson, J., & Hagegård, H. (1995). A singer's expression of emotions in sung performance. In *Vocal fold physiology*. Singular Publishing Group, Inc..

Suzuki, S. (1999). *Suzuki Cello School: Cello Part* (Vol. 3). Alfred Music Publishing.

Suzuki, S. (2003). *Suzuki Cello School: Cello Part* (Vol. 4). Alfred Music Publishing.

Expressive Consistency and Creativity: A Case Study of an Elite Violinist

Stacey Davis[1]
[1]Department of Music, University of Texas at San Antonio, San Antonio, TX, USA

Expert musicians tend to exhibit precise expressive control, thereby allowing them to consistently execute their interpretation across multiple performances of the same piece. However, experts are also distinguished by their creativity, with both deliberate and spontaneous variations occurring in response to the expressive intentions and circumstances of each live performance.

Although research often seeks to identify shared expressive traits across performers relative to musical structure, the study of creativity necessitates the investigation of individual performance characteristics. This project therefore uses a case study approach to explore the relationship between expressive consistency and creativity in performances by renowned Baroque violinist, Rachel Podger.

Podger recorded Bach's solo violin works in 1999 and gave multiple live performances of the same pieces in January 2016. The expressive characteristics of these performances were measured using Sonic Visualiser software, with focus placed on her use of rubato, dynamics, articulation, and ornamentation. Podger also participated in interviews to provide information about the reasons for her deliberate expressive choices, her purposeful changes across different performances, and any spontaneously creative variations that occurred.

Results confirm that Podger created and consistently executes a stylistically appropriate interpretation while also remaining expressively flexible in each performance. The consistency occurs in the location of expressive nuances relative to aspects of musical structure and style (including meter, cadences, contour, voice leading, motivic repetition, and dissonance). In contrast, expressive creativity appears in the use of improvised ornamentation (both melodic embellishments and vibrato) and in the magnitude of timing and dynamic variability both within and across different performances.

The Role Repertoire Choice Has in Shaping the Identity and Functionality of a Chamber Music Ensemble

Alana Blackburn

School of Education, The University of New England, Australia
ablackb6@une.edu.au

ABSTRACT

As we move through the 21st Century, a variety of chamber music ensembles emerge. More unconventional ensembles are appearing, and the rationale of an ensemble's aim or output is becoming more concentrated. No longer can a group be described simply as an 'early music ensemble' or 'new music ensemble', but to find a niche in the ever growing number of small to medium groups, ensembles are becoming more definitive by nature. Through developing a unique approach to performance practice, group identity, and audience perception, the ways in which an ensemble prepares for a performance is determined by a number of aspects. This paper explores repertoire choice as a contributing factor to the successful operation of a chamber music ensemble. Part of a larger study on the processes of chamber music ensembles, this paper focuses on the life experiences of three professional chamber musicians actively involved in chamber music performance, and specifically focuses on the functions and meaning of a group's repertoire choice in the creative and organizational process of a musical group. The study uses a qualitative framework and Interpretative Phenomenological Analysis of three semi-structured interviews as case studies to analyze and describe the meaning of the impact repertoire has on group process, aims and performance. The findings identify the key functions of repertoire choice from a performer and audience perspective including the functionality of the ensemble, verbal and non-verbal communication between players and audience, and the artistic goals and outcomes of the group. The study highlights the sociological importance of the relationship between repertoire and group process, and the need for practicing musicians to fully understand the significance this has on the success of an emerging and/or established ensemble.

I. INTRODUCTION

Despite the variety of genres and styles available to audiences today, classical chamber music ensembles in the 21st century are still traditionally defined as ensemble music '…performed by a small group of players, normally numbering between two and nine, one on a part, usually without conductor…' (McCalla, 2003). The purpose of this paper is to explore how repertoire is used to shape the identity, goals and performance of a classical music group through the personal experiences of professional chamber musicians. Current research explores the professional and social group processes that occur in professional chamber music ensembles with focus on collective decision-making, leadership and performance. These studies have been limited to more 'traditional' or 'conventional' ensembles specifically the string quartet (Butterworth, 1990; Davidson & Good, 2002; Murnighan & Conlon, 1991), wind quintets (Ford & Davidson, 2003), piano duos (Blank & Davidson, 2007) and vocal ensembles (Lim, 2013). This paper fills the gap by exploring the importance of repertoire choice for more unconventional ensembles where typical instrumental roles are not necessarily defined by the musical score.

Discussion of repertoire in studies on group processes in chamber ensembles briefly look at the decision-making processes of how repertoire is chosen, whether it is a group decision or set by an individual (Lim, 2013). Other factors that have been explored are the effect repertoire has on dominance and leadership in terms of musical role of the instrumentalists and how this compares with that of a string quartet both on an artistic level as well as promotion and obtaining work (Blank & Davidson, 2007; Ford & Davidson, 2003). When performing unknown repertoire, types of non-verbal communication through demonstration often exist (Seddon & Biasutti, 2009). Using the string quartet as another example, repertoire choice is also part of the education and training of ensembles, ensuring the canon has been learnt, however, there is little emphasis on the influence of repertoire choice on the aims of a group (Clark, Lisboa, & Williamon, 2014; Cotter-Lockard, 2012). Lim reveals that when choosing repertoire 'efforts were always made to ensure the chosen repertoire would not only reflect the identity of the group, but showcase members' individual talents' (2013, p. 314). Deeper investigation is needed as to how repertoire choice can focus the identity of an ensemble as well as have an effect on the functionality and success of the group.

II. METHODOLOGY

The study takes a qualitative approach employing Interpretative Phenomenological Analysis (IPA) as a research framework. Research in the field of music psychology or performance studies has often taken a quantitative approach, effectively leaving out the subjective and experiential view of the musician (Holmes & Holmes, 2012). As music making or performing is an experience fundamental to a musician and personal to each individual, it is important that the world of the performer is explored and given voice in an ideographical way. Phenomenology is about giving meaning to an experience.

This paper presents three cases as part of a larger study looking at the processes of chamber ensembles and the personal experiences of chamber musicians. Each participant was interviewed in a semi-structured in-depth style for approximately one hour. These interviews explored the themes of identity, the artistic and social factors of working in a chamber ensemble, and the organizational or administrative aspects that make up an ensemble. The participants were able to openly discuss the elements they consider to play the most part in the success of the ensemble they are a core member of.

Each interview was transcribed and then analyzed using Interpretative Phenomenological Analysis techniques set out by Smith, Flowers and Larkin (Smith, Flowers, & Larkin, 2009). Each written transcript was read and re-read in order to make sense of the data and ensure that the focus of the analysis is on the participant while conducting a free textual analysis noting and commenting on the participants' descriptions of things that matter to them and the meaning of those things for the participant (Smith et al., 2009). Emerging themes were then analyzed and any connections across these themes were identified. The IPA method requires that once an in-depth analysis is conducted on one case study, only then can the researcher move on to the next (Smith et al., 2009, p. 82). Analysis was completed using Nvivo software where themes were identified for each individual participant and then common themes were identified across all cases. For these three specific participants repertoire was discussed a number of times, emphasizing importance to the ensemble. Pseudonyms have been used throughout this paper and the ensembles that these participants are members of have not been named. Table 1 provides an overview of the participant case studies used for this paper. It shows the pseudonym used in the study and the professional profile describing the ensemble they are a member of.

Table 1. Participants' background information

Name	Participant Profile
Stella	Repertoire: specializes in 16th and 17th Century polyphony. Homogenous ensemble of 13 members
Brett	Repertoire: not available for this ensemble, arrange and transcribe their own repertoire. Homogenous ensemble of 4 members
John	Repertoire: composer/performer ensemble performing contemporary repertoire. Mixed ensemble of 6 members

III. FINDINGS

A. Case study 1 – Early music ensemble

There is a whole area of musicology devoted to historically informed performance (HIP), which is known for its thorough practice-based research. The research and investigation into earlier periods of repertoire is often the basis on which an ensemble may develop. How this research into historic performance practice shapes an ensemble goes much deeper than specializing in a particular repertoire for Stella and her ensemble. Stella is a core member of an ensemble focused on complex renaissance polyphonic consort music. The repertoire is arranged and transcribed to suit the instrumentation by members of the ensemble. Stella takes pride in researching 16th and 17th century practices and includes these in her ensemble. These practices include performing without a score, performing without a conductor, allowing the musical material to determine who is leading at what point within the piece, and how themes are musically interpreted and played.

On the edge of what normally constitutes a chamber ensemble, Stella's ensemble comprises 13 members. It is a homogenous ensemble (or whole consort in early music terminology) in that members all play the same instrument from the same family. After beginning the ensemble with a conductor, in recent years the ensemble has become leaderless. During this change, the ensemble has embraced its new structure as part of its identity, and Stella believes it reflects a truer display of the repertoire:

...this has become in a way a part of our ensemble identity which is very much related to the concept of consort itself, like we are a group but we don't have one leader, it's all shared as it is in the music we play, the polyphonic music where you don't have one part in the music which is more soloistic but everybody has the same themes, everybody shares the same material and everyone is equally important. So partly this has a very strong influence in the identity of the group. (Stella)

From Stella's description of the group's identity it is evident that the structure of the group, the collaboration and equality of parts is reflected in the repertoire they perform. In terms of leadership, this adds another element to the preparation of the repertoire as the ensemble performs all programs from memory. As Stella explains, this not only requires memorizing the player's individual part, but also to remember who is leading at what point in the piece. Although Stella identifies this as a risk in performance, staying true to consort music material means leadership is shared and the normal hierarchical roles in relation to the order on the score (highest to lowest voice) do not determine leadership.

...it is more of a risk, on the other hand I think it's much more loyal to the music itself because in that kind of music we play, it's not, the top line, ja, it's the highest but not always the most important. The most important parts are actually the middle voices like if there's a cantus firmus it's most likely going to be on the tenor part or even the bass, and if there's imitation it's going to go through all the lines and very often it doesn't start on the top line, so I think musically it makes much more sense to do this divide the leadership and to feel, then also that all these different people in different roles all have a moment of shine or coming, being a little bit in the spotlight in every composition that makes everyone more alert even through it's more challenging. (Stella)

The ensemble does not have a conductor and therefore it is important that leadership is shared within the group. Rehearsal processes require at least one person 'in charge', however, while performing musical decisions are made and led between the players themselves, reflecting the shape and intention of the chosen repertoire.

This way of performing through directing the musical ideas from the members within the ensemble and not in a typical hierarchical way also has an effect on the audience's experience:

...we found it quite interesting to notice that since we don't have a conductor, almost after every concert, if not after each single concert we have given since then, people erm, audience members come to us and say "wow this was so er interesting to watch because I could see how the music was going woo woo [laughs] and how you were looking at each other and some

people were here and some people…" – so even people who don't really know why that happens, or maybe they are not that acquainted with the music they see this interaction and it captures them visually as well. (Stella)

The absence of a conductor has allowed the ensemble to focus on the repertoire in a different way. The analysis undertaken by the group to decide which musical lines introduce new material, the shape in which the music takes, and performing from memory has had an effect on the way this ensemble performs in public; the highest level of non-verbal communication is employed.

B. Case study 2 – Homogenous ensemble

The most widely studied example of a homogenous self-managing ensemble is that of the string quartet, evinced by the research referenced earlier. This genre came to its peak in the 19th century, instilling a working group whereby each instrument and its position within the ensemble dictates its musical role. A genre developed over 250 years, repertoire is abundant, requiring a certain level of fundamental understanding that can be put into practice each time a piece is chosen to be rehearsed and performed. For ensembles that do not have this wealth of repertoire, the identity and shape the ensemble takes becomes centered on the repertoire, needing more commitment and effort on the part of the players to source, commission or transcribe/arrange existing works and adapt them to the ensemble.

Brett is in a homogenous quartet of which there is limited repertoire available. This forces the members of the quartet to arrange music written for other ensembles and instruments, often drawing on existing repertoire from string quartets and percussion ensembles. The ensemble needs to arrange everything and is frustrating for Brett as it can hinder the productivity of the ensemble:

… so every new CD we do we have to arrange all the music. You know it's arranging existing works, but, you know, like, that's a hell of a lot of work to, whenever we even think about doing new pieces, you know, there's a whole stack of work that's got to go into even getting the notes in front of you [laughs]. It's pretty high maintenance [laughs]. (Brett)

To develop and sustain a profile as a chamber ensemble it requires a large commitment from all the members. This is often time consuming as far as rehearsals, preparation, discussion, communication and travel. For Brett, his definition of a 'high profile' ensemble is one of frequency of performance and recording output. Creating their own repertoire takes a large amount of time and therefore restricts the profile of the group:

… it's probably why we don't have a higher profile because, well not only that, but because you can't, you know, like a string quartet or a brass quintet, or whatever recorder quartet, you can't just go to a library and just pic out a rarely performed piece and put that on your program, there's just nothing there. So there's this whole you know which has to happen. So, not a lot of things can happen spontaneously. It all has to be really planned, long-term vision and that's tricky to maintain. (Brett)

As a unique ensemble, they are developing a new genre and identity. The process of creating their own repertoire not only gives the ensemble a sense of fulfillment in performance, but the group is making a contribution to the genre in the hope that it will grow and encourage other instrumentalists to form similar groups.

… that's very problematic for getting more momentum wit the genre. I mean we're establishing the repertoire side of it… (Brett)

Brett understands that this new genre of ensemble means that there is no existing repertoire available. The goal of the ensemble is not only to perform works arranged for the ensemble to an audience, but also to build the genre through making repertoire more readily available. The lack of music requires more forward planning, not allowing for spontaneous or impromptu performances.

C. Case study 3 – Composer/performer ensemble

Composer/performer ensembles are groups that are founded by a composer who also performs within the group playing their own repertoire. In the case of John, he acts as artistic director, composer and performer providing the ensemble with new repertoire as well as searching for unique collaborations seeking to provide context through the works they are performing. The music performed by this ensemble is new, however, John does not label the ensemble as a 'new music ensemble' or 'contemporary music ensemble' as the goals set out by the group do not lend its identity to that of a new music group.

I guess the goal of that initially was to basically play 20th century music generally, mainly, you know, play new works and old works and so forth. But when I say old, still, relatively recent. So it wasn't like a new music ensemble per say it was never really designed to be like a contemporary music ensemble either really. It's interesting like how we became identified later on. (John)

Collaboration and programming are the focus, aims and goals of this ensemble, not specific repertoire from a particular style or genre. Their programs are often mixed with folk, popular or musical theatre. The group identifies with contextualized, interdisciplinary performance.

…we can act-you know and that really changed, and since then like it hasn't really been about the genre or the repertoire, or necessarily the pieces, it's about how the whole thing works as a whole and how the threads kind of interlie and you know, personally I really like other groups that do that which of course there are some, but others you know focus on the pieces, which is great too. (John)

John further explains that through contextualizing and collaborating in programs, the risk of introducing new music to an audience is minimalized:

When we perform new music if you use that as an example, of course there's going to be a variety of responses that audiences have to that, but generally I think because we *contextualize it. It's never really a big issue because it's never music for it's own sake, that risk is not so much there.* (John)

The composer/performer's role is often to produce repertoire, however in this case it is shared amongst the other members of the ensemble depending on their skills and strengths – a true collaborative effort. Because each member is involved in the production of repertoire, and the ensemble is flexible in its formation, this changes the dynamic, leadership and direction of each piece within a program or performance creating a more complex dynamic.

We do a lot of small chamber, so like dos and trios and things like that, and that also means it's different, everyone has a different function. In the past also, less, ah… in different ways now but, in the past there was in terms of like arrangements that were done on sh-, like notes on page, where now a lot of the arrangements have been aural and stuff but notes on pages there were maybe three people that probably contributed equally when it came to doing those arrangements so that also mean that generally those people would lead stuff, because they knew how it would be. That still kind of happens now but in a slightly, slightly different way. (John)

Refusing to be labeled as a 'new music ensemble' despite the currency of the repertoire performed, John prefers to focus on the package or program performed as a whole. The story told through the music and interdisciplinary collaboration is of more importance than the individual pieces, making the performances more 'audience friendly'.

IV. DISCUSSION

In today's cultural environment musicians act as entrepreneurs deciding and searching for their own brand or identity. This includes sourcing their own repertoire and performance opportunities. For the participants presented in this paper, finding repertoire means not playing 'standard' works but sourcing, arranging, and transcribing existing pieces, or commissioning and composing new scores. This process therefore takes time and can effect the momentum of an ensemble, as is the case of Brett's ensemble. By not easily accessing pieces 'off the shelf' timelines are structured around composer availability and members' willingness to assist in searching for repertoire. The chosen repertoire also reflects the goals and desired outcomes of the ensemble. For Stella and John, the repertoire performed is embodied within the artistic vision of the group, representing the values and aims the ensemble identifies with. Brett explains that the artistic output of the group is not only performing or recording repertoire they have arranged, but also building a library for the genre.

The repertoire represented by the ensemble comes with it a number of socio-cultural factors, both within the ensemble and the audience. The knowledge each performer possesses contributes to the musical outcome of the score both on an aesthetic and technical level. Different repertoire is also found to effect the co-ordination of content, allowing the musicians to focus on different aspects in preparation for and in a performance (Davidson & Good, 1997). The ensembles represented in this paper: early music ensemble, homogenous (same instrument family) ensemble, and contemporary eclectic ensemble, the roles of each of the instruments or players is flexible and therefore changes depending on the style of repertoire or each individual piece. This requires a different approach to rehearsal, performance and discussion about interpretation and leadership.

There have been a number of studies exploring the effect known and unknown repertoire has on an audience and part of developing an ensemble today is audience engagement (Ford & Davidson, 2003; Pompe, Tamburri, & Munn, 2013). Knowing *who* you want to play for has an effect on *what* you are going to play for them. John describes his audience as '*a thinking audience*' not '*genre specific*' and that the audience wants '*to be informed of new things and ideas*'. To connect with this audience, John focuses on incorporating elements of history, anthropology and science into his programs, contextualizing the new repertoire the group has created. The musical flow of leadership within Stella's ensemble gives the audience a visual as well as an aural map of the music; the polyphonic writing of the repertoire is shared throughout the ensemble giving new insight into the composer's musical intentions. This results in audiences enjoying music they may be unfamiliar with. The memorization of the repertoire also allows members of the group to engage with and '*focus on the audience*' rather than a conductor or music stand.

Ensemble identity is formed through a number of factors, a title for the group, dress, artistic and organizational roles, and shared goals (Blank & Davidson, 2007, p. 245). The choices behind certain repertoire do shape the values and outcomes of an ensemble but also develop the identity of the group and its *raison d'être*. The aims of an ensemble and why they focus on certain repertoire differ: (i) they do it for purely artistic/musical reasons, (ii) the group is composer driven, (iii) the programs are educational or contextual, or (iv) the group is experimental and interdisciplinary. The repertoire reflects these aims and the three cases presented provide clear examples of well thought out alignment between the group's aims and identity and repertoire choice.

V. CONCLUSION

This paper explored the role of repertoire choice in shaping the social processes and functionality of chamber music ensembles. It highlights the implications of such repertoire choice that in itself is a carefully thought out process that effects the ensemble in terms of sourcing music, leadership, audience engagement, time management and identity. The participants in the three cases presented here identified repertoire as an extremely important factor in how the ensemble is shaped, indicating that in order to maintain a successful ensemble more thought needs to go into aligning the aims, goals and objectives of an ensemble.

Highlighting such an important step in shaping an ensemble means that this process cannot be overlooked when aligning the aims and objectives of a professional ensemble. The information provided by these participants leads to further investigation into audience development and whether this alignment and the impact good repertoire choice has on the audience is worthwhile for an ensemble. This information also informs practitioners and teachers the significance this has on the success of a chamber music ensemble.

REFERENCES

Blank, M., & Davidson, J. W. (2007). An exploration of the effects of musical and social factors in piano duo collaborations. *Psychology of Music, 35*(2), 231-248.

Butterworth, T. (1990). Detroit String Quartet. In J. R. Hackman (Ed.). *Groups that Work (and Those That Don't) Creating Conditions for Effective Teamwork* (pp. 207-224). San Francisco, California: Jossey-Bass Inc.

Clark, T., Lisboa, T., & Williamon, A. (2014). An investigation into musicians' thoughts and perceptions during performance. *Research Studies in Music Education, 36*(1), 19-37

Cotter-Lockard, D. (2012). *Chamber music coaching strategies and rehearsal techniques that enable collaboration.* (Doctor of Philosophy Dissertation). Fielding Graduate University.

Davidson, J. W., & Good, J. (1997). *Social processes in Western Art Chamber Music: Some Aspects of the Social Psychology of the String Quartet*. Paper presented at the Third Triennial ESCOM Conference, University of Uppsala, Sweden.

Davidson, J. W., & Good, J. (2002). Social and Musical Co-Ordination between members of a String Quartet: An Exploratory Study. *Psychology of Music, 30*, 186-201.

Ford, L., & Davidson, J. W. (2003). An Investigation of Members' Roles in Wind Quintets. *Psychology of Music, 31*(1), 53-74.

Holmes, P., & Holmes, C. (2012). The performer's experience: A case for using qualitative (phenomenological) methodologies in music performance research. *Musicae Scientiae, 17*(1), 72-85.

Lim, M. C. (2013). In pursuit of harmony: The social and organisational factors in a professional vocal ensemble. *Psychology of Music, 42*(3), 307-324.

McCalla, J. (2003). *Twentieth-Century Chamber Music* (2nd ed.), New York, NY: Routledge.

Murnighan, K. J., & Conlon, D. E. (1991). The Dynamics of Intense Work Groups: A study of British String Quartets. *Administrative Science Quarterly, 36*(2), 215-288.

Pompe, J., Tamburri, L., & Munn, J. (2013). Symphony Concert Demand: Does Programming Matter? *The Journal of Arts Management, Law and Society, 43*(4), 215-288.

Seddon, F., & Biasutti, M. (2009). A comparison of modes of communication between members of a string quartet and a jazz sextet. *Psychology of Music, 37*(4), 395-415.

Smith, J. A., Flowers, P., & Larkin, M. (2009). *Interpretative Phenomenological Analysis: Theory, Method and Research.* London: SAGE Publications Ltd.

Negotiating incongruent performer goals in musical ensemble performance

Jennifer MacRitchie[1], Andrea Procopio[1], Steffen Herff[1] and Peter Keller[1]
[1]The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Sydney, NSW, Australia

In examining sensorimotor synchronization in an ensemble context where individual goals may differ, we explored how performers negotiate differences in tempo and dynamics instruction in consecutive performances to achieve a successful performance. Differences in synchronization for incongruent/congruent goals (I/C) as well as tempo/dynamics (T/D) instruction were measured. 14 pairs of pianists performed four Bach chorale reductions (each 12 bars) in four conditions (CT,CD, IT,ID), one tempo and dynamics condition for each score. After an initial two bars of steady tempo/dynamic level, instructions were realized as alternating increases/decreases across two bar blocks, repeating over the remaining 10 bars. Incongruent instructions were realized as an increase in tempo/dynamics for one part with a simultaneous decrease in tempo/dynamics for the other part. For each condition, pairs gave four consecutive performances, counted in with a metronome. Pianists sat back to back and performed 'mechanical' readings of the score to obtain a baseline measure of asynchrony. Mean absolute asynchrony values for each condition showed that, as expected, performers synchronized better in congruent vs incongruent instructions, with linear improvements across consecutive performances. Generalised mixed effect models successfully predicted the congruency in the tempo conditions using the correlation coefficient of the each performer's inter-keystroke intervals compared against their "ideal" performance. In the dynamic conditions, the model also showed the correlation between the two performers' key velocities to be significant. Effects of musical expertise and ensemble experience are also discussed. These differences in tempo and dynamics results suggest that ensemble performers will not unilaterally use information regarding each others' actions to coordinate a successful performance.

Limited Evidence for an Evaluation Advantage for Performance Played from Memory

Reinhard Kopiez,[*1] Anna Wolf,[*2] Friedrich Platz[#3]

[*]*Hanover University of Music, Drama and Media, Hanover, Germany*
[#]*State University of Music and Performance, Stuttgart, Germany*
[1]reinhard.kopiez@hmtm-hannover.de, [2]anna.wolf@hmtm-hannover.de, [3]friedrich.platz@mh-stuttgart.de

ABSTRACT

This study investigates the influence of an actual music stand on the evaluation of a videotaped audio-visual solo instrumental performance. Previous research gave evidence that the presence of a score or music stand (obstructing the view of the performer) might negatively influence the evaluation of the performance. However, due to methodological ambiguities, results in previous studies cannot be regarded as explicit. Thus, we conducted a replication study with better control over confounding variables (e.g., varying levels of technical proficiency in different conditions). A violoncello player performed two pieces for solo instrument: one time with a music stand on stage (pretending to play from score) and one time without. The level of technical proficiency was kept constant in both conditions by the use of a pre-recorded well-rehearsed performance from memory. Audio tracks were synchronized with the performance movements in a playback paradigm. Based on the performance evaluations from a web-based experiment (N = 471 participants), we found a significant but small effect size for the main effect of with vs. without a visible music stand (d = 0.27). We conclude that the audience's appreciation of a particular performance by heart might be based on factors other than the objective performance quality.

I. BACKGROUND

Playing from memory has a long tradition in music performance (Chaffin, Demos, & Logan, 2016), and the performance without a visible score or music stand is often regarded as a hallmark of virtuosity. In fact, this is expected from musicians in competitions, examinations and recitals. In a pioneering study, Williamon (1999) investigated the influence of playing from memory on the audience's audio-visual performance evaluation. The author assumed that performance "by heart" makes communication of expressive cues to the audience easier because the performer can move freely and is unconstrained by the visual contact with the music stand. He found that evaluations were biased significantly in favor of performances without a music stand (when the view of the performer was not obstructed). However, neither the experimental design nor the data analysis on the level of single performances and single evaluation items enabled researchers to draw generalizable conclusions.

II. AIMS

In our replication study we took up the work by Williamon (1999) and reviewed his hypothesis that when players perform by heart, they receive a better evaluation from the audience than when they require the aid of a music stand for their performance. In addition, we assumed that the positive influence of playing from memory is more pronounced in musically sophisticated observers (Lehmann & Kopiez, 2013).

III. METHOD

A. Design

A 2 × 2 between subjects design with the factors Performance (playing from memory vs. with the score) and Participant's Degree of Musical Sophistication (high vs. low) was used. The study was conceived as an online experiment. An a priori power analysis by the software G*Power V 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007) with an assumed medium effect size of f = 0.25, α = .05 and 1-β = .90 resulted in a total sample size of at least N = 171 participants.

B. Stimuli

The beginnings of two works for violoncello solo by J. S. Bach (Prelude from BWV 1007 and Gigue from BWV 1009) were selected for the experiment, as they are part of a professional cellist's standard repertoire. Both pieces were well-rehearsed (and memorized) by an advanced music student, and an audio track was recorded in a studio. The cellist obtained both audio tracks one week before the video recordings in order to practice a playback performance to his own recordings.

The video recording was conducted in a concert hall with professional background light. An HD video camera (Canon Legria HF G10) was used as the recording device. The playback track was performed by in-ear monitoring with hidden earphones (masked by camera perspective) and a 2 bar count-in. Video recordings were repeated until the performer and experimenters were satisfied with the synchronization of the video and playback tracks.

The final versions of audio-visual recordings were produced by the software Adobe Premiere (V 5.5), and fade-in/fade-out was added. The stimulus length was limited to about 45 seconds.

C. Paticipants

In total, a convenient sample of N = 471 participants (M_{age} = 28.59, SD = 10.09 years) with 10.3 years (SD = 6.4) of instrumental practice on average took part in the online study.

D. Procedure

Participants gave informed consent. After adjustment of the loudness level and a technical inspection of the internet connection and personal computer audio-visual setup, the experimenter showed a practice video to the participants in order to familiarize them with the rating procedure before they rated the test videos. The entire duration for the study was about 15 minutes, which is within the limits of internet-based experimenting (Reips, 2002).

E. Items for Evaluation

In addition to the items obtained from the original study by Williamon (1999), items from previous performance evaluation studies by Berlo, Lemert and Mertz (1969), Behne (1994), Thompson and Williamon (2003), and Lehmann and Kopiez (2013) were included, resulting in a total of 13 items. A 4-point Likert scale was used for the rating procedure, and participants evaluated each performance by giving marks to the 13 items ("I found the cellist's performance …"; 1 = not at all, 4 = very much).

IV. RESULTS

F. Data Analysis

After we removed n = 94 incomplete data sets, n = 377 cases remained for data analysis. The 13 items were dichotomized and analyzed by Item Response Theory (1 PL model) for Rasch validity. The marks for the 3 remaining items from IRT analysis (*authentic, sonorous, professional*) were summed up and divided by 3, so that the resulting Performance Evaluation Scale (PES) was within a score range from 1 to 4. To increase the contrast between groups of high vs. low level of musical sophistication, we excluded n = 47 participants who were neither musical experts (≧ 8 yrs of music lessons) nor amateurs (1-4 yrs of music lessons). A final sample of n = 330 cases remained for statistical analysis.

G. Differences between main conditions

Figure 1 shows the between groups main effect for the conditions (with vs. without music stand). Evaluations for playing without music stand were slightly more positive (M = 3.20, SD = 0.51) when compared to a performance with a visible music stand (M = 3.05, SD = 0.62). A t-test for this main effect resulted in a significant finding, which was, however, characterized by a small effect size (t(318.47) = -2.47, p = .014, d = 0.27). The absolute difference between both conditions was only 0.15 scale steps (out of a maximal range of 3).

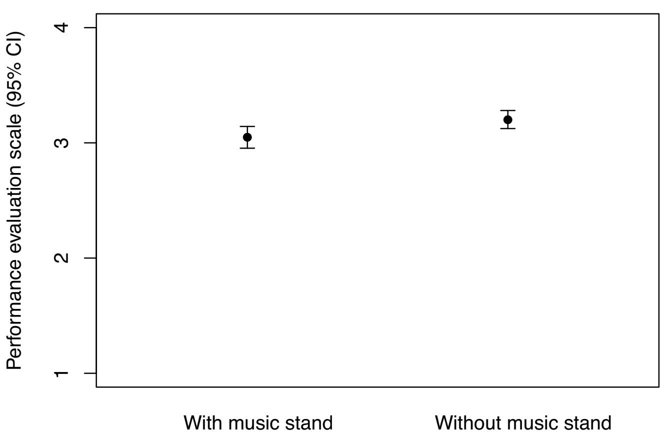


Figure 1. Mean values of the Performance Evaluation Scale for the conditions with vs. without music stand (1 = poor, 4 = very good). In both conditions the performer knew the piece by heart, but in the condition with music stand, he looked at the score.

H. Interaction effects

To test for the hypothesized interaction between presentation mode (with vs. without music stand) and level of musical sophistication (high vs. low), we conducted an omnibus test (2-way between groups ANOVA) for group differences. No significant interaction effects were found for presentation mode × level of sophistication, presentation mode × music piece, and music piece × level of sophistication (all p values > .23).

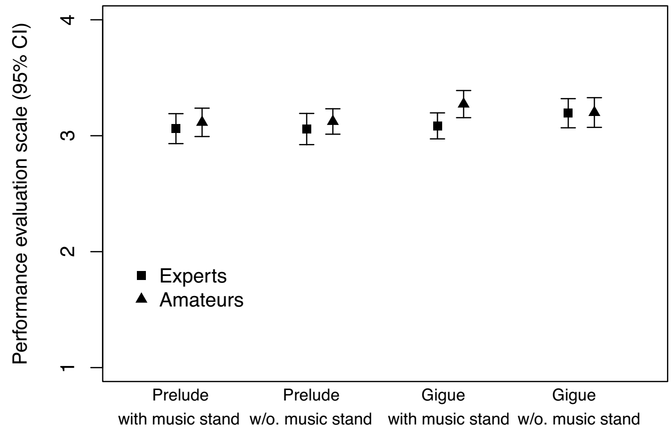


Figure 2. Mean values of the Performance Evaluation Scale for the interaction of with vs. without music stand and high vs. low degree of musical sophistication (1 = poor, 4 = very good). In all conditions, the performer knew the piece by heart, but in the condition with music stand, he or she looked at the score.

V. CONCLUSION

We conclude that the audience's appreciation of a particular performance by heart might be based on factors other than the objective performance quality. Additionally, although performance from the score is often confounded with a player's supposed lack of technical proficiency – which was also the case in Williamon's (1999) study – the mere use of a music stand subsequently does not result in a negative performance evaluation. Against the background of previous research on the complex interplay of variables in evaluation processes, the presumable dominant influence of a single variable on performance evaluation is implausible. In a meta-analysis, Platz and Kopiez (2012) could show that all visual components of stage behavior improved the evaluation of a performance by an effect size of d = 0.51. Moreover, in a study on the features of persuasiveness, Platz and Kopiez (2013) revealed that the formation of the audience's impression is determined by a bundle of factors such as gaze direction, stance width or step size of the performer. According to the authors, the key construct that best describes this complex evaluative process is that of "appropriateness": only in the case of a match between the audience's expectations and the performer's stage behavior will the audience develop the wish for an ongoing performance. As long as it remains unclear as to why we should even feel that a music stand interferes with the performance, then this working tool might have only a marginal influence on our evaluation of a performance.

ACKNOWLEDGMENT

We thank two anonymous reviewers for their helpful comments on an earlier version of this manuscript as well as Cosimo Caloeiro for his patience and cooperation during the recording of the videos and Theresa Tamoszus for her support during the implementation of the study.

REFERENCES

Behne, K.-E. (1994). *Gehört, gedacht, gesehen: zehn Aufsätze zum visuellen, kreativen und theoretischen Umgang mit Musik [Heard, thought, seen: Ten essays on visual, creative, and theoretical dealings with music]*. Regensburg: Bosse.

Berlo, D. K., Lemert, J. B., & Mertz, R. J. (1969). Dimensions for evaluating the acceptability of message sources. *Public Opinion Quarterly, 33*(4), 563-576.

Chaffin, R., Demos, A. P., & Logan, T. (2016). Performing from memory. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (2nd ed., pp. 559-571). Oxford: Oxford University Press.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191.

Lehmann, M., & Kopiez, R. (2013). The influence of on-stage behavior on the subjective evaluation of rock guitar performances. *Musicae Scientiae, 17*(4), 472-494.

Platz, F., & Kopiez, R. (2012). When the eye listens: A meta-analysis of how audio-visual presentation enhances the appreciation of music performance. *Music Perception, 30*(1), 71-83.

Platz, F., & Kopiez, R. (2013). When the first impression counts: Music performers, audience, and the evaluation of stage entrance behavior. *Musicae Scientiae, 17*(2), 167-197.

Reips, U.-D. (2002). Standards for internet-based experimenting. *Experimental Psychology, 49*(4), 243-256.

Thompson, S., & Williamon, A. (2003). Evaluating evaluation: Musical performance assessment as a research tool. *Music Perception, 21*(1), 21-41.

Williamon, A. (1999). The value of performing from memory. *Psychology of Music, 27*(1), 84-95.

Are visual and auditory cues reliable predictors for determining the finalists of a music competition?

Friedrich Platz[1], Reinhard Kopiez[2], Anna Wolf[2], Felix Thiesen[2]
[1]State University of Music and Performing Arts, Stuttgart, Germany
[2]Hanover University of Music, Drama and Media, Hanover, Germany

Judgments of music performance can be explained as the result of a performers' impression formation within a music-related persuasion framework (Lehmann & Kopiez, 2013). Based on this theoretical framework, a recent study by Tsay (2013) has investigated the influence of visual information on the successful prediction of the winner of a piano competition. The author concluded that an above-chance correct identification of music competition winners relied solely on the visual presentation and was not improved in other presentation modes. However, the author's generalizations were based on non-standardized music performances. We therefore conducted a replication study with the aim to investigate the participants' discrimination sensitivity for the correct identification of a music competition's finalists. In contrast to Tsay (2013), but in line with Platz and Kopiez (2012), judgments of an audio-visual music performance were expected to show better evaluations when compared to other presentation modes. Standardized performances of 3 semi-finalists and 3 finalists of an international violin competition were recorded on video. Stimuli were prepared in 3 presentation modes (audio-only, audio-visual and video-only) with a duration of 6 s. In an Internet experiment with a one-way between-subjects design, participants' discrimination sensitivity was measured by means of SDT (Macmillan & Creelman, 2005). A Hierarchical Bayesian approach with multiple comparisons (Kruschke, 2015) was applied to analyze response behavior as a function of presentation mode. Participants in the audio-visual presentation mode showed the highest discrimination sensitivity when compared to the other conditions. Following from that, Tsay's (2013) findings were not replicated. We therefore conclude that the visual component of music performance has been overestimated in recent studies investigating audience's response behavior within strategic decision-making situations.

Music, animacy, and rubato: what makes music sound human?

Adrian S. Blust, David J. Baker, Kaitlin Richard, Daniel Shanahan

***Abstract*—**Our understanding of—and preference for—music is dependent upon the perception of human agency. Listeners often speak of how computer-based performances lack the "soul" of a human performer. At the heart of perceived animacy is causality, which in music might be thought of as rubato, and other variations in timing. This study focuses on the role of variations in microtiming on the perceived animacy of a musical performance. Recent work has shown that the perception of visual animacy is likely categorical, rather than gradual. Although a number of studies have examined auditory animacy, there has been very little research done on whether it might be thought of as a dichotomy of alive/not alive, rather than a continuum. The current study examines the specific intricacies of musical animacy, specifically, how microtiming variations of inter-onset intervals contribute to the perception that a piece was human performed. Additionally, this study aims to examine the possible nature of categorical/continuous perception of musical animacy. In Experiment 1: "Rohum", computer sequenced MIDI renditions were manipulated to contain set random fluctuations of inter-onset intervals. In Experiment 2: "Humbot", participants were presented with human performances digitally recorded using MIDI keyboards, and were asked how "alive" each performance sounded using a 7-point Likert scale. Human performances were divided into ten degrees of quantization strength, increasing from raw performance to 100% quantization. Results suggest an optimal level of quantization strength that is correlated with higher perceived animacy, and fixed random fluctuations of IOI are not a good indicator of human performance. This paper discusses the role of external stylistic assumptions on perceived performances, and also takes into account musical sophistication indices and experience.

***Keywords*—**Animacy, performance, microtiming variations, rubato.

I. INTRODUCTION

THE perception of causality is paramount to the understanding of our surroundings. It is the difference between a falling branch and a stick being thrown at us, between leaves rustling and someone walking behind us. Research in vision has attempted to identify visual cues that are used by the viewer in order to discern whether particular objects are perceived as animate or inanimate. Some of the earliest work on the perception of animacy and causality was conducted by Heider and Simmel who demonstrated the role of motion in the perception of animacy in shapes [1]. Stewart showed that the perception of animacy needs only a few very simple perceptual cues, and experiments by Premack and his colleagues later showed that certain movements allow viewers to ascribe intentions and motives to abstract geometric figures such as triangles and squares [2]–[4]. Tremoulet & Feldman demonstrated that the perception of causality in visual stimuli might be linked to change in direction [5]. Scholl and Tremoulet refer to this as the "energy violation" hypothesis [6]. Broze codified much of this research into six indicators of an animate object [7]:

1. Animate agents are self-propelling or locomotive, and can move under their own power.
2. Animate agents possess intentionality, and exhibit goal-directed behavior.
3. Animate agents are communicative, and employ signalling systems.
4. Animate agents are sentient, and can experience subjective feelings, percepts, and emotions.
5. Animate agents are intelligent, capable of rational thought.
6. Animate agents can be self-conscious, having metacognitive states about their own consciousness and that of others.

It could be argued, however, that these qualities themselves are less important than the perception that an agent has the capacity to convey them. Recent work by Looser and Wheatley has shown that the perception of visual animacy can be understood in agents that lack any locomotive qualities [8]. The study presented photographs showing a sequence of a gradual transition between inanimate and animate faces (dolls and humans, via image manipulation), and asked participants to identify the point where the face "becomes alive". Interestingly, the perception of animacy seemed to be categorical, rather than continuous. Similarly Wheatley, Milleville, and Martin carried out fMRI scans as participants viewed the sequence of morphed photos, and found a change of inferior temporal activation consistent with the animate/inanimate distinction [9] (for similar results, see [10]). While motion, intentionality, or the ability to communicate were not present in these images, given a convincingly animate face, it could be inferred that the image was that of an animate being and thus demonstrate the potential for all of the

A. S. Blust is a recent graduate of the University of Virginia, Department of Cognitive Science & Music, Charlottesville, VA 22903 USA (phone: 202-380-7099; e-mail: asb8nk@virginia.edu)

D. J. Baker is with the School of Music, Louisiana State University, Baton-Rouge, LA, 70802 USA (phone: 414-736-7948; e-mail: Davidjohnbaker1@gmail.com).

K. Richard is with the School of Music, Louisaina State University, Baton-Rouge, LA 70802 USA, (e-mail: richard.kaitlin@yahoo.com)

D. Shanahan is a professor of Music Theory with the School of Music, Louisiana State University, Baton-Rouge, LA 70802 USA, (phone: 614-940-2560; e-mail: daniel.shanahan@gmail.com)

above, rather than demonstrating the explicit ability to do so.

Just as visual cues can be used to infer the animacy of an object, it would follow that auditory cues can contribute to the perception of an animate agent creating a sound. Nielsen, et al. conducted an auditory analogue of [5], using synthesized mosquito sounds [11]. Using binaural spatialization software, the authors varied the direction of motion and velocity of the mosquito sound and asked participants to rate how likely the sound was being produced by an animate source. Interestingly, velocity changes were significantly rated as more animate when compared to the other paradigms of manipulation including directional change of motion or no change at all. It gives rise to the question of what aspects of sound, and its organization in time, can be manipulated to change the perception of animacy in music.

Perhaps the closest analogue to changes of direction, motion, and intentionality in music would be the use of expressive timing. In music, expressivity is regarded by musicians to be the most important aspect of performance characteristics [12], and in performance, emotions are expressed through subtle implicit deviations in timing, dynamics, and intonation. In one study, Gabrielsson and Juslin asked participants to perform a melody to express different emotions such as sadness, happiness, and anger, as well as to perform with "no expression" [13]. The fluctuations in performance timing strongly differed from performances of other emotions. Moreover, timing variations were the most salient cues for an expressive musical performance, and participants that were asked to perform with "no expression" contained the smallest deviations in timing. Juslin defines expressivity as "random variations that reflect human limitations with regard to internal time-keeper variance and motor delays," [14]. Geringer et al. expands upon this, discussing the role of consistent expressivity as necessary for the appreciation of music as human-produced [15].

Manipulations in timing appear to elicit the greatest amount of change in the perception of human character in music. Johnson et al. asked participants to rate their perception of how "musical" a piece of music was (specifically a performance of Bach's Third Cello Suite), and found that when the music was manipulated to contain an exaggerated amount of rubato, "musical" ratings trended downwards [16]. However, as rubato was manipulated to contain less than the original performance, participants perceived the music as significantly less "musical". Bruno Repp analyzed the relationship between microtiming variations and larger scale timing variations found in rubato [17]. When participants were asked to perform Chopin's Preludes No.15 and No.6 with a "normal" expression metronomically (without aid of metronome), and in synchrony with a metronome, Repp discovered that articulated rubato performances of normal expression were simply exaggerated gestures of the random variations of timing that naturally occur due to human limitations. This suggests that the expressive nature of rubato mimics the micro-timing variations that already exist. This provides a stable foundation for the basis of the present study.

This study aims to elucidate the perceptual mechanisms involved in the perception of an animate human performance or an inanimate computer MIDI sequence using micro-timing variations. If microtiming variations are a salient cue for the perception of animacy in music, where is the tipping point between the perception of a "deadpan" computer generated MIDI sequence or an emotionally stimulating human performance? Furthermore, what is the optimal magnitude of variation that induces a convincingly human performance?

Using two experimental paradigms, this study explored how the perception of animacy is affected as a computer MIDI sequence is applied with variance, and from the other direction, how the perception of animacy in a raw human performance changes as microtiming variations are decreased. We hypothesized that animacy in music is most likely not linear, meaning that participants do not hear "aliveness" on a sliding scale, and that there is a likely "sweet spot" in rubato where performances that are too straight will be thought of as robotic, as will pieces that have too much variance.

II. Experiment 1: Applying Variance to Computerized Performances

Our first study used sequenced performances, adding varying levels of variance to each performance. As this study was converting "robotic" performances to more "human" performances, we advertised it as "Rohum" for participants (as a way of keeping it separate from "Humbot", which altered human performances).

A. Methods

Participants. Students of the Louisiana State University (LSU) School of Music (N=38) were recruited for participation (18 Females, 20 Males, mean age: 20.4). Experimental trials were conducted in the Music Cognition and Computation Lab.

Stimuli. Bach's Toccata and Fugue in D minor (BWV565) and Bach's Concerto for Oboe and Violin (BWV1060a) were computer generated using Finale MIDI music notation software (MakeMusic) to produce precisely metrically organized note onsets. Excerpts were split evenly by time (5 seconds each). Using Logic Pro X (Apple Inc.), tempo, timbre, and velocity fluctuations were minimized in order to eliminate any effect of expressiveness outside that of timing. Using Max/MSP software (Version 7; Cycling '74), we applied variance to each recording by randomly adding or subtracting an interval of time determined by dividing a maximum note displacement time of 500ms into 100 equal windows. For example, onsets occurring at 500ms and 1000ms with 5ms variance would be manipulated by randomly adding or subtracting 5ms to create a new manipulated onset of either 495/505ms or 995/1005ms respectively. If the recording were set to include a variance of 10ms, the same excerpt's onsets would be altered so that they fell at either 490/510ms and 990/1010ms after the beginning of the recording.

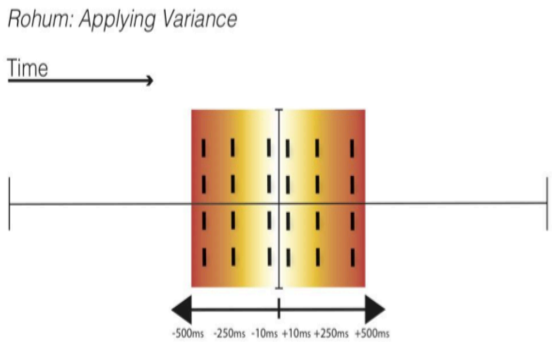


Fig. 1 Experiment 1: Rohum. Variance was applied to each recording by randomly adding or subtracting an interval of time on the order of milliseconds to all note onsets. The amount of variation was therefore fixed and determined by dividing a maximum offset time of 500ms into 100 different degrees of variance increasing at 5ms increments. Centered line represents metric beat; vertical dashed lines represent manipulated note onset times, shifting away from the beat, as variance increases.

Design. Participants were first asked to complete the Goldsmith Musical Sophistication Index, in order to later examine the relationship between ratings and relative levels of musicality [18]. After completing this survey, participants were split into two conditions, each with three blocks of ratings, and each played approximately 50 recordings of MIDI performances (~ 10s long). Participants were told that some of the performances were played by humans and others were sequenced, and were asked "how alive does this performance sound?". Each recording was rated on a 7-point Likert scale (from 1, definitely not alive, to 7, definitely alive).

B. Results

Animacy ratings. The animacy ratings were associated with the degree of variance of each stimulus. As mentioned above, we hypothesized that too much variance would lead to decreased perceptions of aliveness, but that too little variance would evoke a similar response. There is likely an ideal level of fluctuation that would create a sense of animacy. The hypothesis of a significant arc, however, was not supported. As can be seen with in Figures 2 and 3, however, the data fits nicely with a linear model. When fit with a linear regression, there was a significant (negative) correlation between the amount of variance applied and the perception of "aliveness". Results from both BWV565 and BWV1060A yielded significant results ($p<.001$), but with a relatively small effect size.

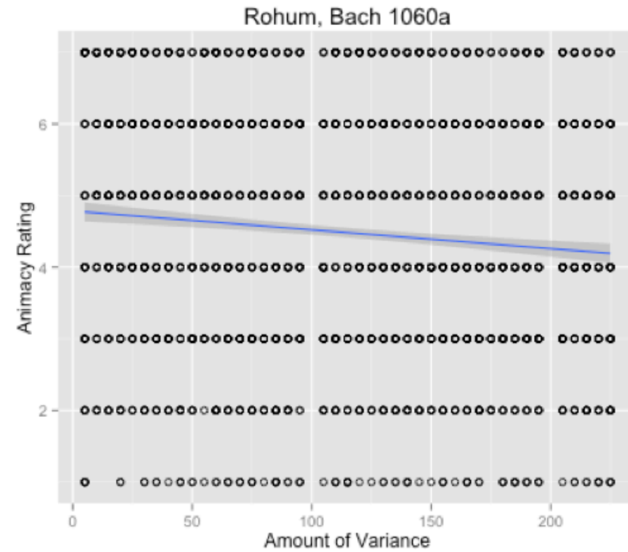


Fig. 2 Linear regression model demonstrating the relationship between the animacy rating of BWV1060a (on the y-axis) and the amount of variance ($p < .001$; small effect size)

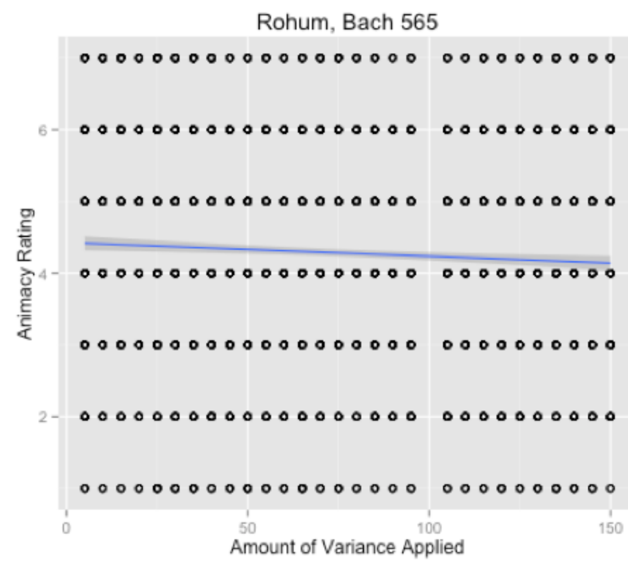


Fig. 3 Linear regression model demonstrating the relationship between the animacy rating of BWV565 (on the y-axis) and the amount of variance ($p < .001$; small effect size)

C. Discussion

The results indicate that even a slight increase in amount of variance makes the music sound more unnatural. The animacy ratings seemed to taper down as variation in inter-onset intervals increased. Increased microtiming variations using these methods did not seem to emulate a real human performance. The results here might also be seen as supporting the findings of Johnson et al. which, as discussed above, found that exaggerated rubato negatively affected the perceived musicality of a performance [16]. In the current study, as rubato increased, "aliveness" went down, which could be seen as a proxy for musicality, in some ways.

III. Experiment 2: Quantizing a Human Performance

The purpose of the second experiment ("Humbot") was to approach the question of how animacy ratings change with manipulation of microtiming variations, but from the opposite direction of the previous experiment ("Rohum"). This study is comprised of both a lab and a web study. The latter study was originally meant as a way of replicating the former study. Ideally, we would see an arc of "animacy ratings", based on the human performances that were quantized to varying degrees.

A. Methods

Stimuli. Humbot. A pianist performed Bach's Concerto for Organ in G major (BWV592) and Chopin's Mazurka 49 in F Minor (Op.68, no.4) on a MIDI keyboard. The pianist was instructed to perform the pieces naturally. MIDI recordings of both performances were divided into 10 degrees of quantization strength. Using Logic Pro X software (Apple inc.), tempo, timbre, and velocity fluctuations were minimized, and quantization strength was systematically applied on a continuum from 0% to 100% quantization by increments of 10%. Both recordings were quantized to the nearest 16th note. Thus, natural variation found in human performance decreased as quantization strength was increased.

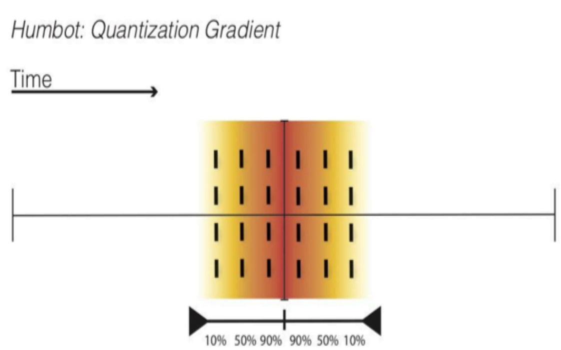


Fig. 4 Experiment 2: Humbot. Natural variance due to human limitations in performance were systematically reduced along a quantization gradient. Recordings were divided into ten different degrees of quantization strength where raw human performance was manipulated to increase in quantization strength from 0% quantization to 100% quantization by increments of 10%. Centered line represents metric beat; vertical dashed lines represent natural human fluctuations of timing in performance, shifting closer to the metric beat, as quantization strength increases.

B. Experiment 2a: Lab Study

Participants. Eighteen participants of the LSU School of Music (10 Males, 8 Females, mean age: 20.6 years, SD= 1.3) were recruited, and were given course credit for their participation. There was an overlap between participants recruited for Experiment 1 and Experiment 2 due to convenience sampling. Experimental trials were conducted in the Music Cognition and Computation Lab at LSU.

Design. Nearly identical to the first experiment, participants were asked to complete the Goldsmith Musical Sophistication Index, and were instructed to listen to approximately 50 recordings (~10-20s long) at degrees of quantization that were randomly selected. Participants rated each recording on a 7-point Likert scale (from 1, definitely not alive, to 7, definitely alive).

C. Experiment 2b: Web Study

Participants. Twenty-nine volunteer participants (14 Males, 15 Females) were recreuited through various social media platforms, such as Facebook, Twitter, and Reddit. Participants were not required to be a part of a university setting, and we did not require previous musical training.

Design. All participants were allowed to access the study from any location. Participants were given a link and voluntarily completed the study using various web browsers. Participants were prompted on the screen to listen to 5-10s recording excerpts at randomly selected degrees of quantization strength. After each stimulus, participants were asked to use the number keys 1 - 7 to rate each recording (from 1, definitely not alive, to 7, definitely alive) before moving on to the next recording. The participant had the choice of how quickly to move through the experiment.

D. Results

Animacy ratings. In order to examine the analogue between visual and auditory animacy, the results were analyzed in a method similar to Looser and Wheatley [8]. Like their study, we analyzed animacy ratings by first transforming them linearly to a scale from 0 to 1 (0 = less likely to performed by a human, 1 = more likely human). In order to find the point at which a recording was equally likely to be perceived as animate or inanimate—this point is known as the point of subjective equality (PSE; [8])— the animacy data from each participant was fit with a cumulative normal function across all recordings. A t-test was then performed between the PSE of each recording and the midpoint of the transform. The results of this t-test were significant ($p < .001$), indicating that the transformation from one end of the "aliveness" spectrum is not linear.

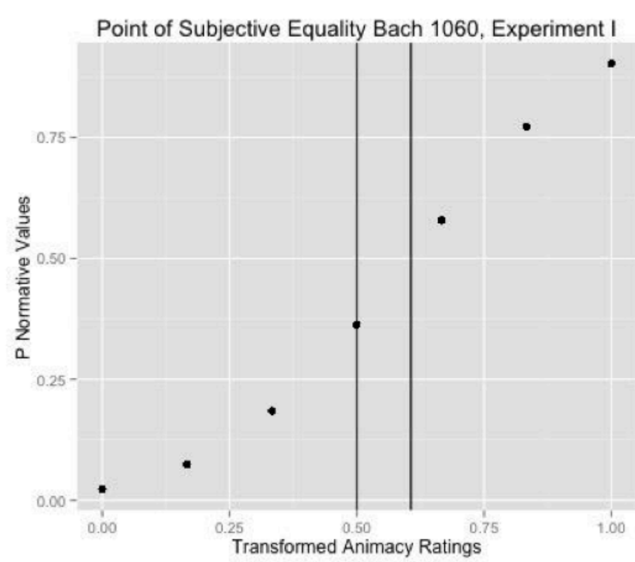


Fig. 5 Experiment 2a, the point of subjective equality for the animacy ratings of Bach's BWV1060a. This indicates that the relationship between the two extremes on the alive/not-alive spectrum do not follow a linear relationship.

E. Discussion

Ideally, we would be able to demonstrate that an arc of "aliveness" exists depending on the level of quantization. We were unable to find such a result with the lab study, which we think might be largely due to the fact that we had a relatively small sample size. As such, we decided to use the web study (discussed below) less as a replication, and more as a way of increasing statistical power.

The web study results did, in fact, indicate a trend toward an optimal level of quantization strength, where highest animacy ratings were weakly correlated. As can be seen in Figure 6 (below), we find that the original performances were correctly rated as "alive", but that these ratings actually increased as a minimal level of quantization was applied. Interestingly, however, these ratings sharply declined as quantization went past a threshold (around 50%). This effect seems to exist most strongly in the Chopin pieces, and less so in Bach, indicating a top-down style-processing component is at play.

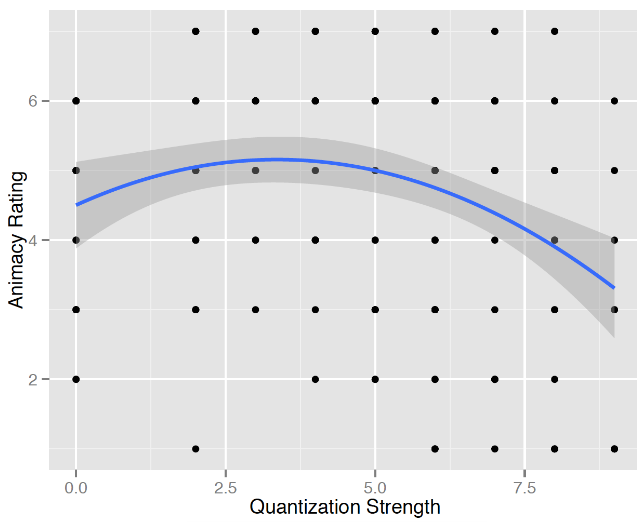


Fig. 6 Chopin's Mazurka 49 in F Minor (Op.68, no.4) from the web-study, indicating that there is likely a point at which there is an optimal level of quantization in rubato. As a corollary to Experiment 1, where too much rubato hindered the perceived animacy, it seems that too much quantization has a similar effect.

IV. Conclusion

In Experiment 1 (Rohum), we hypothesized that increasing variance would result in higher animacy ratings. However, the contrary was found. As variance was introduced, animacy ratings promptly decreased, suggesting that the method of applying variance may have been too rigid to convincingly emulate human performance. Variance was applied by randomly selecting between adding or subtracting an interval of time that was a multiple of 5 in the range of 0 to 500ms. That is, no random jitter was applied within each interval of time, simply, the original note in the music was shifted by a fixed amount of time to randomly fall either before or after the metric beat. Evidence from Experiment 1 suggests that this model for synthesizing natural variation due to human limitations does not communicate a convincing perception of human performance. The unpredictability of random variations makes every song unique, and does contribute to the 'living' character in music [14], but the present study suggests that the perceptual mechanisms involved in discerning between animate/inanimate music are highly sensitive to the type of microtiming variations that are embedded within the music.

Future work will also examine the types of variance and fluctuations that might be applied. Noise and random fluctuations come in a variety of forms, ranging from "white" noise (uncorrelated variance) to "brown" noise (1/f2 noise) [19]. Human performance has been shown to be a mixture of "pink" (1/f noise) and "white" noise [20], [19], [14], [21], [22]. Slower tempos, or larger intervals between beats produces a variance that more corresponds to 1/f noise. A task testing pure reaction time, such as asking participants to press the spacebar as quickly as possible will produce a variance of reaction time intervals that is more associated with white noise [21]. Preference for 1/f noise variance has been found in computer-sequenced music [23], however, the question of how much variance is necessary to produce a convincingly human performance, and if it is a categorical perception, still remains. Experiment 1 of the present study aimed to produce convincing human performances given computer-generated music as a starting point and applying "fixed" variance. Future research would test if pure white and pure 1/f noise produce significantly higher animacy ratings with increased application of variance, compared to the "fixed" variance method. Furthermore, if a PSE is found, employing a same/different task over the PSE would shed light on if the perception of animacy in music is categorical, thus suggesting consistency with the categorical perception of animacy in faces [8].

The results indicate that the manipulation of rubato using variations in microtiming has some effect on perceived animacy. The size of the effect, however, seems to be composer and genre specific. This is consistent with previous literature that suggests we are more likely to perceive objects as being animate than inanimate as evolutionarily, this is best for survival. Interestingly, this perceptual mechanism also extends into the auditory domain as it relates specifically to music perception [8]. The web study results revealed a trend towards an optimal level of quantization strength in the Chopin stimuli where animacy ratings initially increased in an inverse-U shape, reaching an apex point, but then promptly began to drop down as quantization strength continued to increase. We would argue that these minor changes in microtiming serve as a way to display intentionality and are communicative of human performance. Together, these studies illustrate how the perception of animacy in music can be manipulated by systematically adjusting microtiming variations, and perhaps more importantly, how the method of manipulation (Rohum vs. Humbot) affects the perception of animacy. Future work will continue along this train of thought, engaging in more analyses of stylistic differences and the perception of animacy, as well as the specific points at which rubato is applied, and how the shifting of such points affects the perception of aliveness.

REFERENCES

[1] F. Heider, M. Simmel, "An Experimental Study of Apparent Behavior," *The American Journal of Psychology*, vol. 57, pp. 243–259, 1944.

[2] J. A. Stewart, "Perception of animacy," Unpublished Psychology, University of Pennsylvania, 1982.

[3] D. Premack, "Intentionality - how to tell mae west from a crocodile - a review of intentional stance," *Behavioral and Brain Sciences*, vol. 11, 1988.

[4] D. Premack, "Cause/induced motion: Intention/spontaneous motion," *Origins of the Human Brain*, vol. 321, 1995.

[5] P.D. Tremoulet, J. Feldman, "Perception of animacy from the motion of a single object," *Perception,* vol.29, 2000.

[6] B.J. Scholl, P.D. Tremoulet, "Perceptual causality and animacy," *Trends in Cognitive Sciences*, vol. 4, issue 8, 2000.

[7] G.J.Y. Broze III, "Animacy, anthropomimesis, and musical line," Unpublished Graduate Program in Music Theory, The Ohio State University, 2013.

[8] C.E. Looser, T. Wheatley, "The Tipping Point of Animacy: How, When, and Where We Perceive Life in a Face," *Psychological Science*, vol. 21, issue 12, pp. 1854–1862, 2010.

[9] T. Wheatley, S.C. Milleville, A. Martin, "Understanding Animate Agents: Distinct Roles for the Social Network and Mirror System," *Psychological Science*, vol. *18* issue 6, pp. 469–474, 2007.

[10] E. Halgren, P. Baudena, J.M. Clarke, G. Heit, K. Marinkovic, B. Devaux, et al., "Intracerebral potentials to rare target and distractor auditory and visual stimuli. II. medial, lateral and posterior temporal lobe," *Electroencephalography and Clinical Neurophysiology*, vol. 94, issue 4, 1995.

[11] R.H. Nielsen, P. Vuust, M. Wallentin, "Perception of animacy from the motion of a single sound object," *Perception*, vol. 44, issue 2, 2015.

[12] E. Lindström, P.N. Juslin, R. Bresin, A. Williamon, "'Expressivity comes from within your soul': A questionnaire study of music students' perspectives on expressivity." *Research Studies in Music Education*, vol. 20, issue 1, 2003.

[13] A. Gabrielsson, P.N. Juslin, "Emotional expression in music performance: Between the performer's intention and the listener's experience," *Psychology of Music*, vol. 24, issue 1, pp. 68-91, 1996.

[14] P. Juslin, "Five Facets of Musical Expression: A Psychologist's Perspective on Music Performance," *Psychology of Music*, vol. 31, issue 3, pp. 273-302. 2003.

[15] J. M. Geringer, J.K. Sasanfar, "Listener perception of expressivity in collaborative performances containing expressive and unexpressive playing by the pianist," *Journal of Research in Music Education*, vol. 61, issue 2, 2013.

[16] C.M. Johnson, "Effect of rubato magnitude on the perception of musicianship in musical performance," *Journal of Research in Music Education*, vol. 51, issue 2, 2003.

[17] B. Repp, "Relationships Between Performance Timing, Perception of Timing Perturbations, and Perceptual-Motor Synchronisation in Two Chopin Preludes," *Australian Journal of Psychology,* vol. 51, issue 3, pp. 188-203, 1999.

[18] D. Müllensiefen, B. Gingras, J. Musil, L. Stewart, "The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Population," *Plos One*, vol. 9, issue 2, 2014.

[19] P.N. Juslin, A. Friberg, R. Bresin, "Toward a computational model of expression in music performance: The GERM model," *Musicae Scientiae*, vol. 5, 2002.

[20] R. F. Voss, J. Clarke, "1/fnoise in music and speech," *Nature*, vol. 253, 1975.

[21] D.L. Gilden, T. Thornton, M.W. Mallon, "1/f noise in human cognition," *Science*, vol. 267, 1995.

[22] D.L. Gilden, "Cognitive emissions of 1/f noise," *Psychological Review*, vol. 108, issue 1, 2001.

[23] H. Hennig, R. Fleischmann, A. Fredebohm, Y. Hagmayer, J. Nagler, A. Witt, et al., "The nature and perception of fluctuations in human musical rhythms," *Plos One*, vol. 6, issue 10, 2011.

Metaphor use in skilled movement: Music performance, dance and athletics

Jake D. Pietroniro,[*1] Joyce J. de Bruin,[#2] Rebecca S. Schaefer[#3]

[*]*Music Department, University of California, Santa Barbara, CA, USA*

[#]*Leiden Institute for Brain and Cognition / Health, Medical and Neuropsychology Unit, Leiden University, Leiden, The Netherlands*

[1]pietroniro.donato@gmail.com, [2]joyce_de_bruin@hotmail.com, [3]r.s.schaefer@fsw.leidenuniv.nl

ABSTRACT

Metaphors are often applied to skilled movement to convey specific movement instructions. In the current study, musicians, dancers and athletes were asked about how they use metaphors in their daily practice. A mixed quantitative and qualitative approach suggests that metaphor use is widespread among all experts, although the goals differ between groups. Insights into the way metaphors are used in acquiring skilled movement, performance under stress, and communicating about movement may inform pedagogy in all three areas of movement expertise.

I. BACKGROUND

In music performance as well as in dance and athletics, metaphors are thought to be useful in teaching technique (i.e. efficient movement with minimal chance of injury), or to support expressive performance. Although this practice seems to be widespread, we know little about its prevalence among different movement experts, or the specific goals for which metaphors are used. Although several papers on this topic exist within the music pedagogy literature (cf. Barten, 1998; Woody, 2002) others consider metaphors too vague to include in experimental work (cf. Juslin, Karlsson, Lindström, Friberg & Schoonderwaldt, 2006). In the current study we aim to provide support for the notion that metaphors are considered useful, and used often by movement experts. We hypothesize that (a) metaphor use is widespread among all movement experts in individual practice, (b) instructors often use metaphors and encourage students to use them, (c) metaphors are used for a range of goals, specific to the intended outcome of the skilled movement, (d) using metaphors supports teaching, individual practice and performance and that (e) metaphors are useful in collaborating/communicating.

II. METHOD

Using a mixed approach of quantitative and qualitative questions in an online questionnaire, we asked professional musicians, dancers and athletes (n= 27, 15 and 14 respectively) about metaphor use in various aspects of their daily practice. Questions were directed at the different goals of metaphor use (learning something new, practicing, performing, communication/collaboration), the way they were taught or encouraged to use metaphors and their estimation of its usefulness.

III. RESULTS

The quantitative results suggest that metaphor use is fairly widespread among all expert groups, and that it is encouraged by their teachers. Furthermore, metaphor use can indeed serve different goals, such as learning technique, improving performance and controlling nerves. Finally, metaphors were also considered to be useful as communication and collaboration aids. There were no substantial differences between the expert groups, other than that athletes find metaphors useful to control nerves, which was not the case for the musicians and dancers. Further analysis of the qualitative data will likely offer more nuance to these findings.

IV. DISCUSSION

Although the qualitative data will be further elaborated, the initial finding that different movement groups all make use of metaphors and consider them useful supports our main hypothesis. The difference between groups, namely that athletes focus more on controlling nerves, is directly in line with differences in movement goals, where athletic performance differs substantially from expressive performance in music or dance. Further qualitative analysis of the questionnaire will likely shed light on the different ways in which metaphors are a part of the daily practice of movement experts, with implications for expert practice in pedagogy and performance.

REFERENCES

Barten, S. (1998). Speaking of music: The use of motor-affective metaphors in music instruction. *Journal of Aesthetic Education, 32*(2), 89–97.

Juslin, P. N., Karlsson, J., Lindström, E., Friberg, A., & Schoonderwaldt, E. (2006). Play it again with feeling: computer feedback in musical communication of emotions. *Journal of Experimental Psychology: Applied, 12*(2), 79–95.

Woody, R. H. (2002). Emotion, imagery and metaphor in the acquisition of musical performance skill. *Music Education Research, 4*(2), 213–224

Describing Self-talk in Music Practice Cognition

Julie M. Stephens[1], Amy L. Simmons[1]
[1]Sarah and Ernest Butler School of Music, The University of Texas at Austin, Austin, Texas, USA

Music research often examines practice by attending to observable variables (e.g., practice strategy use, time spent). In this study, we take a different approach to describing practice by revealing cognitive processes that underlie musicians' practice behaviors. How does self-talk that underlies practice behavior influence the skill development process? Kahneman (2011) describes self-talk in terms of two systems: System 1 includes impulses and instincts, whereas System 2 monitors intentional ideas and actions. There is some evidence in sports research that these two systems of thinking affect motor performance differentially, and that elite performers use more system 2 thought during practice than do non-elite performers. The purpose of this study was to examine self-talk in the practice of professional and collegiate-level musicians to determine whether proportional differences exist between these two groups in the occurrence of System 1 and 2 thinking. Subjects ($N = 10$) were music majors ($n = 5$) and professional performers ($n = 5$) at UT Austin, UCLA, The California Institute of the Arts, and UC-Davis. Musicians were videotaped for approximately five minutes while practicing one self-identified goal in literature they were currently working on. Musicians were primed that, immediately following practice, they would watch their video recording and provide detailed verbal recall of the thoughts that had occurred during practice. Analyses revealed that, in general, professionals recalled more System 2 thought than did students, although there is some variation to consider. There were notable differences in the *nature* of the statements made by professionals as compared with those made by university students. Clearly, there are limitations to consider given the small sample size. However, these data illuminate aspects of music practice cognition that may inform how musicians monitor and optimize their thinking during practice.

Outliers in Performed Loudness Transitions: An Analysis of Chopin Mazurka Recordings

Katerina Kosta,[*1] Oscar F. Bandtlow,[#2] Elaine Chew[*3]

[*]*Electronic Engineering and Computer Science, Queen Mary University of London, U.K.*

[#]*School of Mathematical Science, Queen Mary University of London, U.K.*

[1]katerina.kosta@qmul.ac.uk, [2]o.bandtlow@qmul.ac.uk, [3]elaine.chew@qmul.ac.uk

ABSTRACT

In this study, we investigate the relationship between dynamic markings in the score, such as ***p*** (piano), and performed loudness. In particular, we examine the performed loudness for consecutive pairs of distinct dynamic markings. The main aim of this study is to understand expressive variations in a musician's playing style relative to other musicians by analysing outliers in the ways performers navigate transitions between dynamic markings.

The focus of the study is recordings of forty-four Chopin Mazurkas, each having an average of 45 recordings. The dynamic (loudness) values (in sones) are normalized and discretized so as to obtain comparable values derived from different recording environments. This paper describes two experiments: the first deals with outliers in interpretations of dynamic changes from one marking to the next; the second investigates unusual interpretations of dynamic changes for entire recordings. A marking pair is considered an outlier if the value of its corresponding dynamic change from one marking to the next lies in the top or bottom quartile of the distribution of the dynamic change values. A recording is considered an outlier if it is grouped into a cluster having the least number of recordings for that piece.

We find that the largest proportion of outliers appear at the marking pairs (***p***, ***f***), (***f***, ***p***), (***pp***, ***f***), and (***pp***, ***p***). Cortot has the highest number of pairwise-transition and overall recording outliers. Mazurka Op. 24 No. 2 has the highest proportion of marking-pair outliers. Finally, we find that some pianists with high numbers of outlying recordings tend to co-occur in numerous outlier clusters, meaning that they diverge from common practice in similar ways.

I. INTRODUCTION

Varying of loudness is one of the most important expressive devices in music performance. Other devices include articulation, rhythm, and intonation (Burkhart, 1983). A performer communicates the composer's intentions by following the music score, although Cook writes, "according to Schenker, composers' scores represent the effects they wish to create, not the manner in which these effects are to be created" (Cook, 2013, pg. 37). One outcome is that the diversity in various performances of the same piece creates a variety of different patterns when considering their general dynamic behavior (Kosta, 2014).

We are interested in studying the relationship between dynamic markings in the score such as ***p*** (*piano*, meaning soft) and ***f*** (*forte*, meaning loud), and actual loudness as realized in performance. While dynamic markings are important indications of how pieces should be interpreted, their meaning depends on the context in which they appear. In this study, we investigate dynamic (loudness) differences among a large number of recordings of Chopin's Mazurkas and how performers navigate transitions from one dynamic marking to the next.

The main aim of this study is to understand expressive variation in playing styles by analysing dynamics changes at transitions between markings. More specifically, we have conducted two experiments, the first one explores unusual (sometimes contrarian) interpretations of dynamic changes from one marking to the next, and the second investigates unusual interpretations of dynamic changes through entire performances.

This study is not the first to explore extreme performances. An example in case is a detailed explanation of Cortot's unique playing style in his recording of Chopin's Berceuse, Op. 57 interpretation (Wilkinson, 2015). In addition to conveying the composer's intentions, performers also bring to bear their own personal, cultural, and historical viewpoints in subjective expressions (Fabian, 2014).

The remainder of this paper is organized as follows: In Section II we describe the creation of the dataset for this study; in Section III we present the description and the results from the two experiments; finally in Section IV we present the conclusions.

II. DATA AQUISITION

In this study we used audio data from the Mazurka Project (CHARM, 2004). The data consists of 2000 recordings of forty-four Chopin Mazurkas; details of the data set and numbers of recordings are documented in Table A in Appendix A. For this dataset, ***pp*** occurs 62 times, ***p*** 234, ***mf*** 21, ***f*** 170, and ***ff*** 43. The total number of markings is 530. From the audio recordings, we obtain the performed loudness at each dynamic marking in sones. The loudness is smoothed by local regression using weighted linear least squares and a second degree polynomial model (the "loess" method of the MATLAB function *smooth*.)

To determine the loudness values for each beat, the audio has been score-beat aligned. We used as score the Paderewski, Bronarski and Turczynski edition. For each Mazurka, the first author manually annotated the beats for one recording, and beat positions were transferred automatically to the remaining recordings using a multiple performance alignment heuristic. The alignment uses Ewert et al's algorithm (Ewert, 2012), which is based on Dynamic Time Warping using chroma features; the heuristic optimises the choice of reference recording (more details in Kosta, 2016.)

Dynamic values (in sones) have been normalized in order to make comparable values derived from different recording environments and audio engineer mixing choices. More specifically, if $\{y_n\} \in \mathbb{R}$ is the sequence of loudness values (in sones) for each score beat indexed $n \in \mathbb{N}$ in one piece, then the normalised sequence is defined as

$$y'_n = \frac{y_n - y_{min}}{y_{max} - y_{min}}, \quad (1)$$

where y_{min} and y_{max} are the minimum and maximum loudness values, respectively, in the recording. The loudness value associated with a marking at beat b is

$$\ell_b = \frac{1}{3}(y_b + y_{b+1} + y_{b+2}). \quad (2)$$

By the time we have the dynamic value that corresponds to each marking, we create our data in a pairwise manner by subtracting from $\ell_b$the value of the previous marking, ℓ_{b-1}. The result is discretised as follows:

$$f(x) = \begin{cases} 2, & x > 0.5 \\ 1, & 0 < x \leq 0.5 \\ 0, & x = 0 \\ -1, & -0.5 \leq x < 0 \\ -2, & x < -0.5 \end{cases}$$

where the input $x = \ell_b - \ell_{b-1}$ for every marking position b in each Mazurka. The methodology relies on the linearity of the sone values so that we are able to create levels of a same linear distance. This discretization steps allows us to handle the noisy loudness data.

III. RESULTS

Two experiments examine the changes in dynamics in the various interpretations reflected in the recordings of the Mazurkas. The first experiment explores unusual interpretations of loudness changes across marking pairs. The second experiment investigates unusual loudness interpretations by considering dynamic changes across entire performances.

A. Experiment I

We compare dynamic levels at different pairs of markings that appear in the scores. i.e., we consider consecutive pairs (ℓ_{b-1}, ℓ_b), which represent the transition from the marking at position *b – 1* to that at position *b*. A marking pair is an outlier if its corresponding dynamic change lies in the upper or lower quartile of such changes. Thus, outliers represent unusual changes in dynamics from a particular marking to the following one, for example getting softer in a transition where in most recordings the music gets louder.

The results indicate that the highest proportion of outliers appear at the (***p***, ***f***), (***f***, ***p***), (***pp***, ***f***), and (***pp***, ***p***) transitions. We observe two ways for outliers to occur: the first is to overshoot or undershoot the change from one marking to the next in relation with the most common trends; the second is to contradict the change from one marking to the next in opposition to the most common behavior.

Fig. 1 shows how the outliers are distributed, giving their proportions for every marking-pair over all Mazurkas where the specific transition occurs. Note that no data points exist for the pairs (***mf***, ***ff***), (***mf***, ***pp***), and (***ff***, ***mf***) as these marking pairs do not appear in the Mazurkas we analyze.

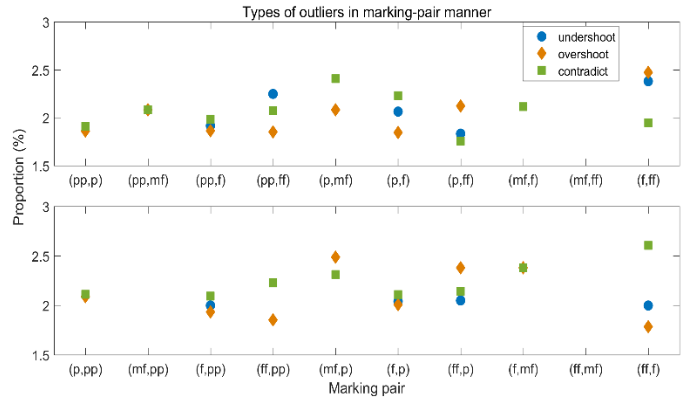


Figure 1. Distribution of outlier transitions as a proportion of all transitions between each dynamic marking pair (circle=undershoot, rhombus=overshoot, square=contradictory.)

A transition that includes many outliers in the (***f***, ***p***) at the beginning of Mazurka Op. 59 No. 3. This transition had a mean value of -1, but over fifty-six recordings eleven recordings overshot to -2, ten recordings made a contradictory +1 transition, and one recording even had an extreme contradictory transition of +2. The score where the specific markings are located is presented in Fig. A in the Section Appendix B. It is worth mentioning that the specific pair has the biggest portion of outliers.

If we consider only pianists that have recorded more than thirty (out of the forty-four) Mazurkas, Cortot is the pianist with the highest number of outliers, as shown in Table 2.

Table 2. Pianists whose recordings had the highest proportion of outlier transitions.

Pianist	# Outliers	# Recordings	# Pairs	Ratio
Cortot	74	42	472	0.1568
Magin	43	35	379	0.1135
Rangell	47	37	416	0.1130

Table 3 presents the recordings with the highest rate of outliers. Rubinstein's recording of Mazurka Op. 30 No. 1, Ashkenazy's Op. 59 No. 2, and Block's Op. 68 No. 3 appear on the list.

Table 3. Recordings having the highest proportion of outlier transitions.

Mazurka	Pianist (year of recording)	# Outliers	# Pairs	Ratio
M30-1	Rubinstein (1952)	5	7	0.714
M59-2	Ashkenazy (1999)	5	7	0.714
M68-3	Block (1995)	5	7	0.714

Mazurka Op. 24 No. 2 has the highest proportion of outliers. Fig. 2 displays the dynamic values for all recordings of Mazurka Op. 24 No. 2; out outlying transitions are highlighted with solid lines. In some cases, like the transition from the second to the third marking, there are no outliers despite the diversity of the dynamic behaviour. The reason is that two distinct sub-populations of the values have been created at the specific cases.

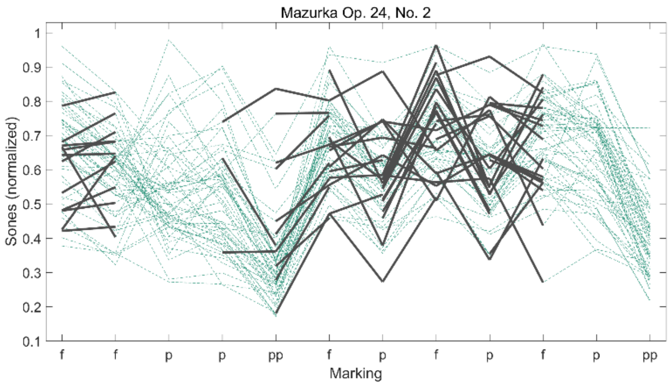


Figure 2. Loudness graphs for performances of Mazurka Op. 24 No. 2 (the one having the highest proportion of outliers); highlighted in bold are outlier transitions.

B. Experiment II

One question that follows naturally from Experiment I is just how varied are the loudness behavior in the recordings? In order to investigate this question, we map each recording to a time series of discretized values of the function f defined in Section II. Each recording then becomes a curve with discretized points in score time. The purpose of the procedure above is to create meaningful clusters of curves so as to further analyze the resulting shapes, and distinguish unusual behaviors.

To cluster the curves obtained, we implemented the k-means machine learning method. The number of clusters, k, is defined by the "gap statistic" (Tibshirani, 2001); we further limit k to the range [3,8] to ensure a reasonable number of recordings appear in each cluster, but also provide the flexibility of detecting clusters that include curves that follow an unusual behavior at the same time.

The result of the process described above is a number of loudness behaviour clusters for each Mazurka. Clusters with the least number of recordings are labeled as outliers; then, we create a list of pianists that appear at these outlier clusters. Table 4 presents the top three outlier cluster pianists who have recorded more than thirty Mazurkas.

Table 4. Pianists having the highest proportion of recordings in outlier clusters.

Pianist (year of recording)	# Outlier clusters	# Recordings	Ratio
Cortot (1951)	12	42	0.2727
Poblocka (1999)	9	33	0.2727
Sztompka(1959)	8	38	0.2105

Magin is the only pianist whose recordings were the single element which was contained in a cluster at the cases of Mazurka Op. 7 No. 1 and Mazurka Op. 33 No. 3. Fig. 3 highlights the loudness behavior of Magin's recordings of both Mazurkas, and how they differ from the other recordings..Fig. 3 shows the centroid for each cluster, thus the discretized dynamic value may not be equal to an outcome of the function f.

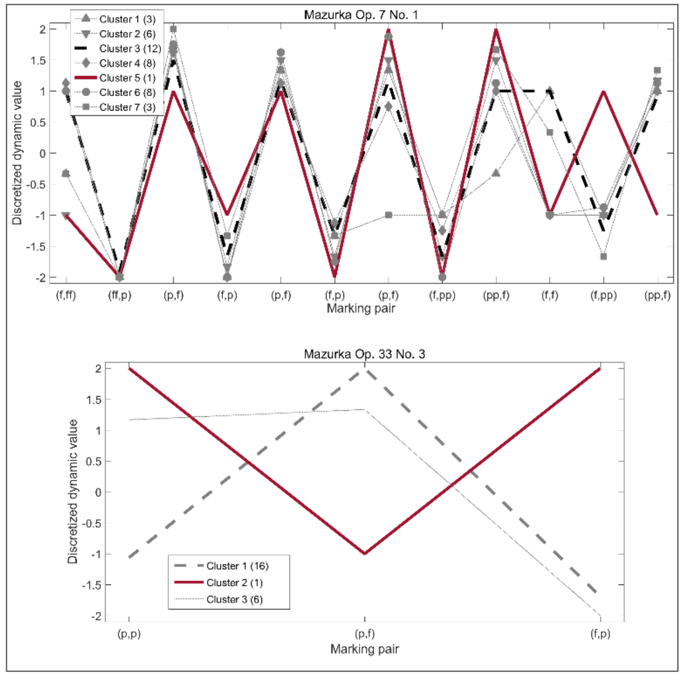


Figure 3. Loudness transition clusters found for Mazurka Op. 7 No. 1 (top) and Op. 33 No. 3 (bottom); in each case, the cluster comprising of only Magin's recording is shown in solid bold lines, the cluster having the highest number of recordings is shown as dotted bold lines.

Next, we focus on when there are commonalities in unusual interpretations. For this, we consider pianists whose recordings are in outlier clusters for every Mazurka, and we compute the number of times two pianists' recordings are classified into the same outlier cluster for all Mazurkas. The results are shown in Table 5, where we give the number of outlier clusters containing the specific pianist pairs. We focus on pairs that co-occur in outlier clusters for more than twenty Mazurkas.

Table 5. Pianists whose recordings co-occur in more than twenty outlier clusters.

Pianist	Fliere	Milkina	Ezaki	Barbosa
Chiu	0	21	0	21
Smith	27	0	0	0
Rubinstein	0	26	0	0
Kushner	0	0	21	0
Czerny-Stefanska	0	0	0	21

Table 5 shows higher degrees of similarity in performed loudness between pianists Chiu—Milkina, Chiu—Barbosa, Smith—Fliere, Rubinstein—Milkina, Kushner—Ezaki, and Czerny-Stefanska—Barbosa. Thus, although certain pianists are more often in outlier clusters, the ways in which they differ in their loudness interpretations often follow shared patterns.

IV. CONCLUSIONS-DISCUSSION

In this study we have analyzed outliers in loudness interpretations in piano recordings. The results show that we are able to distinguish specific dynamic marking pairs which have a high proportion of outliers. In addition, we are able to detect recordings for which the interpretations do not follow the emerging patterns. There is considerable agreement between an interpretation and information derived from a score, nevertheless it is a matter of the pianists' musical choices to control a sense of longer-term intensity modulation.

The reason why an expressive variation happens is "due to deliberate expressive strategies, music structure, motor noise, imprecision of the performer, or even measurement errors" (Langner, 2003.) In any case, the different choices followed by the pianists hold the key to a performance stimulated by creativity and imagination.

ACKNOWLEDGMENT

This research has been funded in part by a Queen Mary University of London Principal's studentship.

REFERENCES

Burkhart, C. (1983). Schenker's theory of levels and musical performance. In David Beach, ed., Aspects of Schenkerian Theory. New Haven: Yale University Press, 95-112.

Cook, N. (2013). Beyond the Score: Music as Performance. Oxford University Press.

Ewert, S., Müller M., Grosche, P. (2009). High resolution audio synchronization using chroma onset features. *Proceedings of IEEE International Conference of Acoustics, Speech, and Signal Processing (ICASSP)* (pp.1869-1872). Taipei, Taiwan.

Fabian, D., Timmers, R., Schubert, E. (2014). Expressiveness in Music Performance: Empirical Approaches across Styles and Cultures. pp. 60-61. Oxford University Press.

Kosta, K., Bandtlow, O. F., Chew, E. (2014). Practical implications of dynamic markings in the score: Is piano always piano? *Proceedings of the 53th Conference on Semantic Audio*. London, U.K.

Kosta, K., Bandtlow, O. F., Chew, E. (2016). Mapping between dynamic markings and performed loudness: a machine learning approach. *To appear in Journal of Mathematics and Music,* special issue on Machine learning and Music, July.

Langner, J., & Goebl, W. (2003). Visualizing expressive performance in tempo-loudness space. *Computer Music Journal,* 27(4), 69-83.

Leech-Wilkinson, D. (2015). Cortot's Berceuse. Music Analysis, 34(3), 335-363.

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a dataset via the gap statistic. Journal of the Royal Statistical Society B, 63:2, 411-423.

V. APPENDIX A

Table A. Chopin Mazurkas used in this study, the number of recordings for each one, and the number of dynamic markings that appear in each one. Mazurkas are indexed as “M<*opus*>-<*number*>.”

Mazurka index	M06-1	M06-2	M06-3	M07-1	M07-2	M07-3	M17-1	M17-2	M17-3	M17-4	M24-1
# recordings	34	42	42	41	35	58	45	50	36	67	46
# markings	18	13	22	13	13	18	7	6	9	7	4

Mazurka index	M24-2	M24-3	M24-4	M30-1	M30-2	M30-3	M30-4	M33-1	M33-2	M33-3	M33-4
# recordings	56	39	54	45	50	54	55	48	50	23	63
# markings	12	7	33	8	14	25	18	5	16	4	12

Mazurka index	M41-1	M41-2	M41-3	M41-4	M50-1	M50-2	M50-3	M56-1	M56-2	M56-3	M59-1
# recordings	35	42	39	33	45	40	67	34	48	51	41
# markings	12	5	6	7	15	14	17	14	7	16	8

Mazurka index	M59-2	M59-3	M63-1	M63-3	M67-1	M67-2	M67-3	M67-4	M68-1	M68-2	M68-3
# recordings	56	56	42	62	35	31	40	42	38	48	42
# markings	8	11	9	4	18	10	13	11	12	21	8

VI. APPENDIX B

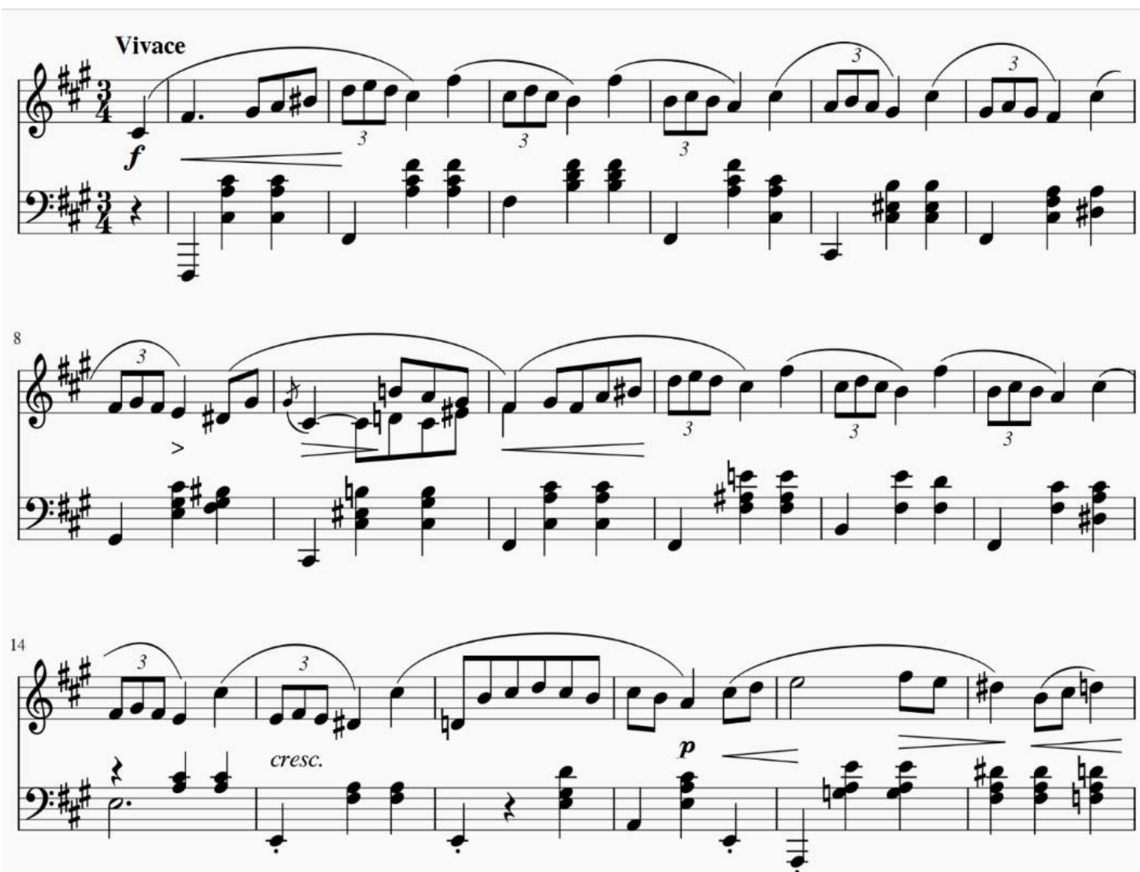



Figure A. Score excerpt from Mazurka Op. 59 No. 3 where the markings of the first pair (*f*, *p*) appears. This transition has the highest proportion of outliers.

How Stable are Motor Pathways in Beginning Wind Instrument Study?

Laura A. Stambaugh[1]
[1]Department of Music, Georgia Southern University, Statesboro, GA USA

A musician has many choices of where to direct her attention while playing. The field of motor control has identified these possible areas as "focus of attention" (FOA). FOA can be directed internally (embouchure or fingers) or externally (keys or the sound of one's music). Previous research in motor learning and one limited music study found an external FOA generally leads to more efficient and effective movement than an internal FOA. However, a recent study indicated university novice woodwind players were not differentially affected by internal and external FOAs, while the advanced players performed most evenly and accurately using an internal FOA.

The purpose of the current study was to examine the effect of FOA on second year band students' performance. It built directly on the previous study, using a repeated measures design with a control condition (no directed FOA), an internal focus condition (fingers), and an external focus condition (sound).

Participants were forty band students in their second year of study, aged 13-14 years (n = 16, woodwinds: flute, clarinet, and saxophone; n = 16, valved brass: trumpet and French horn; n = 8, trombones). The study stimuli were isochronous, alternating two pitch patterns (e.g., eighth notes C-A-C-A-C-A-C). Students heard a model recording of each stimulus at a specified tempo and were then directed to play the measure as evenly and accurately as possible (control condition), while "thinking about your fingers" (internal focus condition), and while "thinking about your sound" (external focus condition). Participants played eight trials of each stimulus. The design was fully counterbalanced, with the exception that the control condition was always performed first. Approximately 24 hours after the first study session, participants returned for a retention test, playing each stimulus three times with no directed FOAs. Results will be analyzed as within and between group ANOVAs for pitch accuracy and pitch evenness.

Performers' motions reflect their intention to express local or global structure in melody

Madeline Huberth[1], Takako Fujioka[1]
[1]Department of Music, Center for Computer Research in Music and Acoustics, Stanford University, USA

When playing music, musicians' body movements relate not only to playing their instrument directly, but to their expressive intentions and interpretation of musical structure. These expressive gestures have been shown to correspond with the level of the phrase, especially in performers' head and torso. A phrase though, can consist of multiple melodic groups. As such, a performer can choose to either express the local groups, or more strongly integrate them into a whole phrase. These choices have been less explored with relation to performer motion. This study aimed to characterize the nature of motions associated with either choice. We filmed 13 cellists playing a musical excerpt, the opening 8 bars of the 3rd Ricercar by Domenico Gabrielli, in two conditions: in which they were asked to think either about local melodic groupings or a whole phrase, prescribed by the experimenters. Cellists were further required to use the prescribed fingerings and bowings, and memorized both interpretations. Cellists were filmed by two video cameras (front and right side perspective), and 2 successful takes of each version were collected. 1-inch squares of retroreflective tape were placed on the forehead, chin, right cheek, hands, shoulders, and torso. Marker positions (in 2D pixel space) for each camera were subsequently extracted using computer vision techniques. Our results show that participants' heads move more in the local task than in the whole-phrase task, measured by distance traveled by forehead and cheek markers. This result is consistent with previous studies relation between head motion and phrase, and extends the relation to show that melodic interpretation is also shown in body movement. The indication of expressive melodic intentions by musicians' bodies has implications for the extent of the embodiment of musical structure in performers.

Cognitive and affective empathic responses to contemporary western solo piano performance

Mary C. Broughton[1] and Jessie Dimmick[2]
[1]School of Music, The University of Queensland, Brisbane, Queensland, Australia
[2]School of Psychology, The University of Queensland, Brisbane, Queensland, Australia

Effective audience engagement with music performing arts involves cognitive and affective elements. Musical and motor expertise, and modality of presentation may shape cognitive, but not affective responses to music performance. In an empathy framework, we investigate cognitive and affective responses to contemporary solo piano performance presented in audio and audio-visual formats. *Affective empathy* involves feeling and sharing emotions (basic attributes of normal-functioning humans). *Cognitive empathy* involves taking the perspective of another, and may be expertise moderated. Understanding arguably results from the cognitive empathic ability to 'put oneself in another's shoes'. Participants were non-musicians (Nms), musician non-pianists (Ms), and musician pianists (Mps). Cognitive (*understanding* – 'put yourself in the performer's shoes' and report your understanding of what the performer is doing to physically generate the expressive performance) and affective (arousal and valence) empathic responses were made continuously. Liking and familiarity ratings were made on seven-point Likert scales after each excerpt. As hypothesized: understanding ratings were higher for Ms and Mps than Nms; understanding and liking ratings for audio-visual presentations were higher than audio-only; liking ratings were higher for Ms and Mps than Nms; no significant arousal, valence, or familiarity effects were observed. Contrary to expectation, understanding ratings did not significantly differ between the two musician groups. Affective empathic responses appear similar across expertise groups and presentation modality. Cognitive empathic responses vary according to musical, but not piano-playing motor expertise, and presentation modality. Potential sample size and cognitive empathy measure issues also warrant consideration. Non-musician audiences may be developed for contemporary music performance through multimodal educational strategies to increase understanding of performance.

Silent Reading and Aural Models in Pianists' Mental Practice

Nina Loimusalo*[1], Erkki Huovinen*[2]

*Department of Music, University of Jyväskylä, Finland
[1]nina.j.loimusalo@student.jyu.fi

ABSTRACT

This study addresses musicians' learning outcomes and subjective experiences in two common types of mental practice: silent score reading and score reading while listening to the music. The study incorporates expert ratings of performances before and after mental practice, questionnaire data concerning modal preferences, as well as semi-structured interviews. The results revealed individual differences in learning outcomes, attitudes toward the two types of mental practice, and the use of imagery. The participants' attitudes and strategies were variously affected by their ability to audiate newly encountered music, their possible preference for learning by ear, and their need to process the score at their own leisure. The results suggest that different types of mental practice might usefully serve various individual purposes in instrumental education.

I. INTRODUCTION

Mental practice, or imaginary practice, recently also called mental simulation, was once defined by Richardson (1967) as "the symbolic rehearsal of a physical activity in the absence of any gross muscular movement" (in Driskell, Copper & Moran 1994). In a meta-analysis of relevant literature, Driskell, Copper, and Moran (1994) found it important to distinguish between *mental practice*, in a strict sense of the term, and *mental preparation* in a broader sense, encompassing also various psyching-up techniques. Recently, van Meer and Theunissen (2009, 94) described *mental simulation* as "a technique by which the mind creates a mental representation of a preconceived idea or action with the intent to practice in order to enhance performance." Mental practice has proved to be effective both for cognitive and physical tasks, but greater efficacy has been found for tasks that included more cognitive elements (Driskell, Copper and Moran 1994; van Meer & Theunissen 2009). It is not to be understood as a substitute for actual physical practice with an instrument, but as an additional practice technique—more beneficial than no practice at all, albeit by itself less effective than physical practice. As a topic that was traditionally researched in the field of sports psychology, mental practice has grown over the years to be an interesting research area in many distinct fields, such as stroke rehabilitation, surgery education, and music. Common interests among the fields have been learning and enhancing performance.

The research on musicians' mental practice has concentrated mostly on learning, memorization, and instrument-related features. Already in an early study with pianists in the 1940s, overlearning by mental practice was found to be superior to overlearning by physical practice in a musical memorization task (Rubin-Rabson 1941). In several studies of musicians, the combination of mental and physical practice has shown benefits equal to those obtained by the same amount of physical practice alone (Ross 1985; Coffman, 1990; Theiler & Lippman 1995; Miksza 2005).

Musicians' mental practice might consist merely in thinking about playing the music in one's mind, but it could also incorporate activities such as reading a notated score and/or listening to a recording of the work. Comparing the use of an aural model with silent mental practice has shown some contradictory findings. Coffman (1990), in a study with pianists, found silent mental practice and mental practice with an aural model to be equally effective. Lim and Lippman (1991) researched mental practice with and without an aural model in memorization of piano music, and found mental practice together with heard music to be less effective than physical practice, but superior to silent score reading. In his study within elementary wind instrumentalists, Fortney (1992) found an aural model combined with score reading superior to silent practice, as well as to free physical practice conditions. Theiler and Lippman (1995) studied mental practice with guitarists and vocalists, and found an aural model to improve several dimensions of performance (especially with vocalists), suggesting that instrument related features might also play a role for the effectiveness of mental coding modalities. Also the level of expertise seems to affect practice outcomes and the need for sensory input: Palmer and Meyer (2000) found a stronger motor independence in mental plans for expert musicians than for novices. Cahn (2008), in a study on learning jazz chord progressions, found task difficulty to affect mental practice achievements and suggested that the superiority of physical to mental practice may hinge on task difficulty.

Aural skills have been found to be an important factor for memorization of music. Highben and Palmer (2004) found strong aural skills and auditory forms of mental practice to be beneficial for memorization. This suggests the need for studies comparing various mental practice conditions within individual performers. Bernardi, Schories, Jabusch, Colombo and Altenmüller (2013) studied individual differences in strategies of mental practice and found optimal memorization associated with pitch imagery and a more general habit for formal analysis. Recently, Bernardi, De Buglio, Trimarchi, Chielli, and Bricolo (2013) found mental practice to improve movement velocity, timing, and coordination in musical performance; movement velocity was improved by motor imagery during mental practice, but hindered by structural analysis of the music.

Individual differences may, according to van Meer and Theunissen (2009), affect the successfulness of mental practice. The present study was developed to shed light on individual differences in and behind skillful pianists' mental practice of new repertoire. We focused on two common mental practice conditions: silent score reading and score reading combined with an aural model of the music.

II. METHOD

A. Participants

The participants were 23 skilled piano students (incl. 22 piano majors) at three Finnish conservatories or universities, with a mean age of 25.3 years (*SD* = 3.6), and and average of 15.0 years of active piano playing (*SD* = 5.4). 17 of them were females. All of the participants had taken "instrumental examination D" in classical piano performance, often understood as an admission requirement for university-level instrumental studies in the country. The participants had participated in professional music studies for an average of 4.5 years (*SD* = 2.6). They reported spending 15.8 hours a week in active music-making (*SD* = 11.2), and 14.5 hours in reading music notation (*SD* = 10.2), on average.

B. Musical materials

Two lesser-known piano pieces by the Lithuanian composer Jurgis Gaižauskas (1922–2009) were selected as musical materials for the study. The two target pieces, *Pute vejas* and *Oi mergele tu jaunoji*, both taken from the same suite of folk songs for the piano, were of comparable length, sharing a roughly similar linear two-voice texture, *andante* character, and broadly modal, but "modern" characteristics in the pitch structure. All written information and all dynamic markings were removed from the scores in order to match the amount of information available. (For better comparability in terms of musical and notational complexity, we further took the liberty to implement one clef change in *Pute vejas* as well as two time signature changes in *Oi mergele tu jaunoji*, from which we also removed a briefly appearing middle voice.)

C. Procedure

The participants were tested individually by a research assistant who also collected their eye-movements during the musical tasks (to be analyzed in a later publication). The participant was seated at a Roland FP-7 electric piano, with a computer screen behind the keyboard for displaying the written music and the instructions. After getting used to the setup, the participant was asked to sight-read one of the target pieces that appeared on the screen. A suggested tempo (the same as used for the corresponding aural model) was given with a few metronome clicks after which the participant performed the piece *prima vista*. Following this initial sight-reading performance, there was a mental practice period of 2'55'', followed by another performance of the same target piece. Finally, this progression of *prima vista* performance—mental practice—*seconda vista* performance was repeated for the second target piece, but now with a different set of instructions for the mental practice phase. All of the performances were recorded using Logic Pro X 10.2 software.

The participants were randomly assigned to four groups of a 2 x 2 design in which roughly equal numbers of them were given each of Gaižauskas' pieces as the first target piece (followed by the other one), and in which roughly equal numbers of participants also began the session with one of two mental practice conditions. In the *aural model* condition, the participants spent the practice time by reading the score displayed on the screen while listening to an aural model of the music (an expressively "flat" performance played by Sibelius 7 software), repeated for three times without a pause. In the *silent reading* condition, the participants spent the same amount of time studying the score in silence.

In accordance with our interest in strategic choices in musicians' freely conceived mental practice routines (cf. Bernardi et al. 2013), the only restriction imposed on mental practice was that the participants were not allowed to touch the keyboard during either of the practice phases. Hence, in particular, such motoric strategies as finger tapping or even "playing in the lap" were *not* discouraged. In general, how "mental practice" would be understood was left open in the instructions.

After the performances, semi-structured interviews were conducted in order to elucidate the relative benefits of the two conditions, and to shed light on potential individual differences in the use of aural imagery. During the interview, the participants were asked to compare the two conditions, explaining their potential preference for one of them. In addition, they were asked whether and to what extent they had been able to "hear the music in their heads."

After the interviews we administered a Finnish version of VARK (7.8) questionnaire of learning styles (Fleming 2012), which distinguishes between visual, auditive, reading/writing, and kinesthetic modal preferences (© Copyright Version 7.8 [2014] held by VARK Learn Limited, Christchurch, New Zealand).

D. Data analysis

The technical level of performance before and after mental practice was estimated by calculating pitch errors and rhythmic errors. The total sum of pitch errors (incorrectly played notes, extra notes and missing notes) was proportioned to the total number of notes in each piece. Each quarter-note beat that included a rhythmic error was counted as one rhythmic error. The sum of rhythmically incorrect beats was proportioned to the total number of beats in each piece.

Learning outcomes were assessed by an expert panel of four skilled pianists and piano pedagogues. The panelists evaluated the improvement during mental practice by comparing the performances before and after mental practice on 7-point scales (between –3 and 3) in terms of technical fluency, expressive interpretation, structuring of the phrases and structuring of the overall form.

III. RESULTS

In the following, we will first introduce our general quantitative results, and then proceed to the more specific qualitative findings concerning individual differences and preferences.

E. Learning outcomes and effects of modal preferences

Table 1 presents the learning outcomes—the average changes in expert ratings between the two performances, and the corresponding changes in error scores. In terms of the expert panel ratings, no significant benefits were found between the silent condition and the aural model condition.

However, a comparison between the pitch errors made in the two performances of each target piece revealed a significant difference between the silent and audio conditions. In the silent condition, mental practice diminished the amount of pitch errors made by 9.9 percentage points, which was significantly better than in the aural model condition.

Table 1. Learning outcomes in the silent and aural model conditions.

		silent	aural	Difference
Improvement	technical fluency	.75	.70	t(22) = .280 *p* = .782
	expressive interpretation	.96	.93	t(22) = .124, *p* = .902
	structuring of the phrases	1.05	.89	t(22) = .753 *p* = .459
	structuring of the overall form	.92	.90	t(22) = .131 *p* = .897
pitch errors		–9.9	–3.4	t(22) = –2.17 *p* = .041
rhythmic errors		–3.4	–3.1	t(22) = –.105 p = .917

The improvement scores, as given by the expert judges, were found to be significantly correlated with scores from the VARK questionnaire for modality preferences especially for the silent condition. In the silent condition (see Table 2), the scores for the kinesthetic and auditive dimensions were significantly correlated with the learning outcomes assessed by the expert panel. In particular, kinesthetic VARK scores found significant correlations with the scores for improvement in expressive interpretation, improvement in the structuring of the phrases, and improvement in structuring of the overall form. In addition, the auditive VARK scores were correlated with the improvement in expressive interpretation.

Table 2. Silent condition: Pearson correlations between learning outcomes and modality preferences

		VARK kinesthetic	VARK auditive
Improvement	technical fluency	.371	–.241
	expressive interpretation	.596**	.540**
	structuring of the phrases	.567**	.163
	structuring of the overall form	.506*	.248
pitch errors		–.362	.260
rhythmic errors		–.317	.124

** p < .01, * p<.05

In the aural model condition, the only significant correlation was found between the visual VARK dimension and the improvement in structuring of the overall form (r = .464, p < .05).

F. Musicians' experience of the aural model condition

Above, we saw that the silent reading and the aural model conditions were equally efficient in facilitating the musicians' learning processes on the group level. Of course, this does not rule out the possibility that individual musicians might find one of these conditions as more suitable, useful, or motivating for themselves. In this and the following section, we will give an overview of the most prevalent kinds of argument that the participants used in the interviews after the experiment to account for their experience of the advantages and disadvantages of the two conditions.

One of the most important benefits of the aural model, as experienced by the participants, was that such a model could offer them a "better general impression of the piece" (Eva; similarly Ella, Anna-Liisa, Anja, Sissi, Tomi, Hilma, Tia). Hearing the music would help the musician "discern everything more clearly" (Tomi), or "perceive and internalize the music better" (Ella; similarly Hilma). Heard sound might, indeed, be felt as a kind of glue between the visual and the kinesthetic:

> The aural model kind of helped to connect the visual score and the melody to doing it with your hands. There were so many aspects that were linked together. (Ella.)

The participants who stated their preference for the aural model typically explained this by their *inability to clearly hear music directly in their minds* on the first encounter (Tomi, Eva, Ella, Sissi, Ville, Anja, Anna-Liisa, Sanna, Anu). Indeed, Tomi suggests that such inability might be due to the very use of audio recordings in practicing:

> I grasp the idea of the music more effectively when I listen to a recording. My inner ear is not that developed—probably because I have not practiced it. Maybe I could develop it if I left out the recordings. (Tomi.)

The aural model, then, was found helpful for "confirm[ing] the right pitches" (Esko; similarly Tomi, Anu, Hilma), and for detecting possible mistakes in the previous *prima vista* performance (Milla). Some of the participants reported "being more self-confident" after hearing the music (Tia; similarly Tomi, Mari).

It was especially in connection with the aural model that the participants also mentioned reflecting on *expressive aspects of performance*. Hearing the melody was found to invite such reflections more than the silent condition did (Sissi, Ella, Anna-Liisa, Eva, Esko). Heard sound might be experienced to free some additional resources for thinking about the dynamics (Ella, Anna-Liisa), or to "remind of the articulation" (Eva). Important elements to be emphasized such as slurs would "come forth better" while listening, "reminding [the player] of performing them" (Eva). Occasionally, the details perceived with the support of the heard music might be devoted more prolonged visual attention, even while the music was progressing onward:

> [With the aural model] I was thinking about interpretational things, how I would play it. […] For example, when I heard some passage, I thought: hey, I could play like this and this. […] When the theme came for the second time, I stopped and thought to myself how it might be played differently than in the beginning. (Sissi.)

The heard audio version may also have been used as such as a guiding model for expressive performance (Eva, Anna):

> I noticed that in the latter piece I took quite a lot from the heard aural image, so that I found a more vertical treatment to the piece. (Eva.)

More often, however, the deadpan computer performance seems to have been treated critically, "encouraging a more musical performance" in exchange (Tomi). Interestingly, some of these participants reported "fighting in my mind with the computer" (Kaisa; similarly Hilma), or comparing the heard sound with a better imaginary version (Esko, Ella). Esko, who confessed "not thinking much about the interpretation" in the silent condition, did so with the audio condition:

> I was maybe comparing my interpretation to what I was hearing. […] For example, some notes are supposed to be a little bit less in volume than others. In such instances, I noticed that the computer is just playing it [in] the same volume, and I was comparing it to some ideal interpretation.

Even though the aural condition often complemented participants' weaker audiation skills, some participants with strong audiation skills might also prefer the aural condition as a tool for *learning music by ear*. Marika and Mari both reported having perfect or near-perfect pitch, but confessed to be slow music readers, and hence they valued the auditory model. Mari calls herself an "auditive learner," and recounts:

> With a new piece, hearing it helps me a lot […] Maybe it's because as a musician one of my weaknesses is handling the rhythms. (Mari.)

Learning by ear was a characteristic approach to new repertoire for Mari, who actually scored highest of all participants on the VARK auditive modality scale. She mentioned "reading the score being challenging since childhood." This is how she described her exceptional memory for music as a combination of absolute pitch and visual imagery:

> In general, being with the score has always been such that I play a lot without it—like from the memory and by ear […] because I can play a lot based just on an auditory image, so that I don't need a score […] [The memorizing] is very easy for me. Often just hearing it once will allow me to play it [...] Of course, I immediately hear the key and how it begins, and I can locate where I'm at [in the music]. […] It happens somehow automatically. In my opinion I don't consciously think much of anything—rather, I just listen, and then I'm like: "a-ha: this is how it goes," and then I go and play it […] I can't explain it very well. […] I kind of hear, and I see—I can very well imagine fingers on the keyboard: "a-ha, that's the way it goes here, and now it does this." (Mari.)

The participants also pointed out some disadvantages with the aural model. Chief among them was a tendency to *inhibit slower and more detailed processing* of passages that would have required extra time for analysis or simulated practice (Mai, Tuomas, Anna, Matti, Susanna). Tuomas found the aural model "superficial," and thus experienced it as challenging for himself to "balance between" listening to the model and focusing on musical detail (Tuomas). Mai who wanted to "look at the notes at their places at my own leisure" simply found the tempo too fast and was "frustrated by it": as she said, "my brain just could not keep up that speed" (Mai).

Some participants found the mere presence of heard sound *distracting* to mental practice (Tuomas, Sami, Kaisa). Tuomas seemed to conceive of the "practicing" element in mental practice as something separate from either reading the score or listening to the music, and found it "too much" to combine all of these aspects:

> I hardly did much practicing with the music playing in the background—I could not concentrate. I could have done that too, but it would have equaled doing like three things at the same time: looking at the score, physically practicing and listening to the music. (Tuomas.)

Interestingly, some of the participants even claimed that hearing the aural model actually *prevented them from mentally listening* to the music (Tuomas, Anna, Mai, Susanna). Susanna actually tried to ignore the aural model, concentrating instead on visual and kinesthetic strategies:

> I had to go like: "Okay, close your ears and concentrate on the score." I was kind of able to exclude [the aural model] reasonably well, but […] because I also wanted to hear it in my mind—to get an aural picture of it, but then I was forced just to look at the score, and think about how it would feel on the keyboard. It was disturbing not to be able to listen to it in my head. (Susanna.)

G. Musicians' experience of the silent condition

The most often reported benefit of the silent condition was the possibility it offered for *processing the score at one's leisure* (Anu, Anna, Matti, Susanna, Tuomas, Mai, Mari, Tia). In a word, this condition gave the musician "more time to practice in one's mind" (Anu). Compared to the aural model condition, more varied reading processes were reported here. These ranged from simply reading through the music (Anu, Esko, Ella, Sanna, Anna, Mai), through more segmented (Tuomas, Susanna) and comparative reading processes (Susanna, Esko, Matti) to "unsystematic" (Anja), or "squinting" (Tomi) approaches.

The silent condition was found helpful in aiding *detailed analytical study* of important or difficult passages (Susanna, Tuomas, Anna, Milla). Susanna, who in this condition received the highest expert judgments across the board, enjoyed the condition, as it enabled her to quickly scan the score for difficulties and to concentrate on them:

> With the silent [condition], I was able to focus my concentration—like "Okay, this is easy, easy, that is a difficult one, difficult one, that's a difficult one." And then, to concentrate on those. To go through them many times. (Susanna.)

Another potential benefit of silent mental practice is that, for some musicians, it might actually *facilitate audiating the music*. One indicator for playing the music in one's mind, as opposed to analytical scrutiny, might be tapping to the silently imagined music (as described by Tia and Hilma). Without support from heard music, enacting the music kinesthetically might be experienced as a helpful approach:

> I was moving my fingers quite a lot. I moved them actually more than with the music playing [in the aural model]. It was somehow even more important to get the feeling image with it. (Tia.)

Many of the participants who expressed preference for the silent condition reported having easily heard the musical score in their minds (Anna, Esko, Susanna, Kaisa, Sami, Milla)—a skill commonly referred to as *audiation* (Gordon 1984). Anna who described getting the music "more carefully in my head" in the silent condition, noted that playing the music in her mind helped her "pretty much with everything," also helping her to relax. Descriptions of audiation in the silent condition involved such phrases as "beautiful melody" (Esko) or "getting the atmosphere" (Anna), suggesting that at least some participants were able to grasp the silently audiated materials as aesthetically rich *music*.

Whereas the aural condition was above noted to evoke thoughts concerning music's expressive aspects, the silent condition sometimes seemed to render "thinking about the interpretation more difficult" (Ella, similarly Sissi). At least these two participants were among the ones who also admitted not having been able to audiate the music clearly.

In cases where audiating the pitch content of the piece was found to be challenging, it could also be replaced with audiating the rhythms (Sanna, Matti). This could be understood as a kind of silent singing:

> I could not somehow sight read it by singing in my head. I didn't want to spend time in searching for the pitches or the pitch intervals. Maybe I was thinking more about the rhythm. […] I was singing it in my head even though it was just the rhythms at this point. In this piece, I did not [think about] the melody. [Sanna demonstrates by playing on the lap and singing in a whispery tone:] Dii-dii-dii. Somehow I was in that pulse. (Sanna.)

Some participants found the silent condition "useless" (Tomi) and were even pondering "when it is going to end" (Anu). The disadvantages of the silent condition were very often related to the participant's inability to clearly listen to the music in one's mind (Ella, Eva, Sanna, Sissi). Ella mentioned feeling a certain "insecurity" when she "did not know if it's precisely correct."

IV. CONCLUSION

The present study addressed the role of individual differences in musicians' mental practice of new musical material, meaning score-based practice away from the musical instrument. Two different mental practice conditions—silent mental practice and mental practice while listening to the music—were compared through expert ratings of achieved improvement as well as through interviews of the participants.

The two conditions did not differ in their overall benefits, as judged by the expert panelists. However, the learning outcomes in the silent condition were associated with kinesthetic and auditive tendencies, as assessed by a test of modality preferences. In the silent condition, the expert panel scores for improvement in expressive interpretation, structuring of the phrases, and structuring of the overall form were significantly correlated with a kinesthetic tendency; this was not the case for improvement in technical fluency. These results resonate with previous findings suggesting that mental practice might be more efficient for cognitive than for motor tasks (see Driskell & al. 1994; van Meer & Theunissen 2009). Now, as found by Kessler, Rubinstein, Ginsborg, and Henik (2008), a cued manual motor imagery is often present during music reading. We suggest that when processing the score in silence, individuals with stronger kinesthetic tendencies might more fluently have been able to translate cognized musical elements into embodied action.

In the silent condition, the connection between auditive tendencies and improvement in expressive interpretation probably cannot be accounted for by conscious planning of expressivity: in this condition, only very few of the pianists reported such planning during the silent practice. One potential explanation for the connection between auditive tendencies and expressive improvement could simply be that an "auditive musician" is better able to expressively respond to her/his own playing in the course of performance. An alternative explanation would be the use of audiation strategies during mental practice. Namely, in the silent condition, the four participants who improved their expressive performance the most reported singing or listening to the music in their minds during silent practice; none of the four participants with least improvement in expressivity reported clearly hearing the music. Having internalized the piece through inner audiation, the musician not only memorizes the piece better (Highben & Palmer 2004; Bernardi & al. 2013), but also gains more freedom to even improvise a cogent expressive interpretation in the actual performance.

In the aural model condition, the expert panel scores for improvement in structuring of the overall form were associated with a visual tendency. Many musicians experienced that the aural model offered them a better general impression of the piece. We suggest that the aural model may have helped especially musicians with visual tendencies to connect the visual score and the heard music to project a more structured understanding of the overall form.

As revealed by the interviews, musicians' ability to hear newly encountered music in their minds affects their attitudes towards types of mental practice. Silent mental practice may often be found useful by musicians with developed audiation skills. When audiation skills are more modest, listening to a recording may be found necessary for grasping a more holistic impression of the piece. However, listening to an aural model may also be useful for musicians with preference to "learn by ear," irrespective of their level of audiation skills. However, the habit of using recordings to speed up the learning process in the beginning stages of learning new music may turn out to be an impediment to the development of stronger skills in score reading and audiation.

Of the two mental practice conditions considered, it was especially the aural model condition that evoked thoughts about expressive interpretation—paradoxically, despite the expressively flat computer rendering used here for the aural models. While some of the musicians accepted the performance on the recording as a guiding model, others would, conversely, be inspired to their own expressive performances in opposition to the deadpan version heard. Aural models may surely be used as sources of inspiration for expressive interpretation. Nevertheless, they may sometimes guide performers too much, preventing them from creating an expressive interpretation of their own—which was noted by some pianists as a negative side of using recordings in early stages of learning.

The aural model, despite giving support, may also disturb the cognitive processing by preventing the musician from processing the piece in a slower and more detailed manner. Some of our participants also reported that the aural model prevented them from audiating the music by themselves. Silent score reading gives more freedom to process at one's own leisure, to concentrate, to focus on critical passages, and to engage in structural analysis. It can be seen as a more ideal type of mental practice, particularly in situations that require more in-depth, detailed processing of the music. It is, perhaps, a more ideal activity for developing skills like audiation, reflective score analysis or imagined physical action.

In silent mental practice situations, musicians are not exclusively dependent on their ability to audiate. If one's ability to silently "listen" to the music from the score is weaker, the skills lacking in aural imagery may be compensated for by audiating only the rhythms or leaning on kinesthetic strategies such as tapping in rhythm. Despite individual weaknesses, there is always something a musician can do. The challenge is to find the most appropriate strategies for different individuals and situations.

In our experience, instrumental educators do often advise their students in mental practice, but judging from the characteristically individual attitudes of the musicians in our study, it is a real possibility that not all of the methods suit everybody. Our findings converge with previous research in emphasizing the importance of audiation for mental practice, but perhaps meaningful programs of mental practice could also be developed along other lines—for instance, by building on an individual's kinesthetic tendencies. Our study tentatively suggests that individuals' kinesthetic tendencies may play a bigger role in the silent processing of music than heretofore acknowledged.

ACKNOWLEDGMENTS

This research was supported by the project *Reading Music: Eye Movements and the Development of Expertise*, funded by the Finnish Academy (project nr. 275929).

REFERENCES

Bernardi, N. F., De Buglio, M., Trimarchi, P. D., Chielli, A., & Bricolo, E. (2013). Mental practice promotes motor anticipation: Evidence from skilled music performance. *Frontiers in Human Neuroscience*, *7*(451), 1–14.

Bernardi, N. F., Schories, A., Jabusch, H-C., Colombo, B., & Altenmüller, E. (2013). Mental practice in music memorization: An ecological-empirical study. *Music Perception*, *30*(3), 275–290.

Cahn, D. (2008). The effects of varying ratios of physical and mental practice, and task difficulty on performance of a tonal pattern. *Psychology of Music*, *36*(2), 179–191.

Brodsky, W., Kessler, Y. Rubinstein, B.-S., Ginsborg, J., & Henik, A. (2008). The mental representation of music notation: Notational audiation. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(2), 427–445.

Coffman, D. (1990). Effects of mental practice, physical practice, and knowledge of results on piano performance. *Journal of Research in Music Education*, *38*(3), 187–196.

Driskell, J. E., Copper, C., Moran, A. (1991). Does mental practice enhance performance? *Journal of Applied Psychology*, *79*(4), 481–492.

Fortney, P. M. (1992). The effect of modeling and silent analysis on the performance effectiveness of advanced elementary instrumentalists. *Research Perspectives in Music Education*, *3*, 18–21.

Gordon, E. E. (1984). *Learning sequences in music: Skill, content, and patterns*. Chicago, IL: G. I. A. Publications, Inc.

Highben, Z. & Palmer, C. (2004). Effects of auditory and motor mental practice in memorized piano performance. *Bulletin of the Council for Research in Music Education*, *159*, 58–65.

Van Meer, J. P. & Theunissen, N. C. M. (2009). Prospective educational applications of mental simulation: A meta-review. *Educational Psychology Review*, *21*(2), 93–112.

Miksza, P. (2005). The effect of mental practice on the performance achievement of high school trombonists. *Contributions to Music Education*, *32*(1), 75–93.

Palmer, C. & Meyer, R. K. (2000). Conceptual and motor learning in music performance. *Psychological Science*, *11*(1), 63–68.

Ross, S. L. (1985). The effectiveness of mental practice in improving the performance of college trombonists. *Journal of Research in Music Education*, *33*(4), 221–230.

Rubin-Rabson, G. (1941). Studies in the psychology of memorizing piano music. VI: A comparison of two forms of mental rehearsal and keyboard overlearning. *Journal of Educational Psychology*, *32*(8), 593–602.

Theiler, A. M. & Lippman L. G. (1995). Effects of mental practice and modeling on guitar and vocal performance. *The Journal of General Psychology*, *122*(4), 329–343.

Auditory and motor learning for skilled and novice performers

Rachel M. Brown[1] and Virginia Penhune[2]
[1]Department of Neurology and Psychopharmacology, Maastricht University, Maastricht, The Netherlands
[2]Department of Psychology, Concordia University, Montreal, Quebec, Canada

Acquiring sensorimotor skill, such as playing the piano, requires learning patterns of actions and sounds. The role of actions and sounds in sensorimotor skill acquisition may depend on expertise. Skilled pianists have developed strong action-sound associations, in which highly-practiced action patterns can be generalized to produce multiple possible sound patterns. Pianists also engage motor brain regions when simply perceiving related sounds, suggesting auditory-to-motor neural connections. Therefore, pianists may learn to play new melodies most efficiently by listening, while novice performers may learn most efficiently through motor practice. We examined whether skilled pianists learn to perform new melodies more efficiently, and show greater motor brain response, when listening compared to performing, in contrast to novices. Pianists and non-musicians underwent functional MRI scanning while learning novel piano melodies in four conditions: by following visuospatial cues (cue-only learning), or by additionally listening (auditory learning), performing (motor learning), or performing with auditory feedback (auditory-motor learning). Participants performed melodies from memory (recall) periodically during learning. Pianists' pitch accuracy was highest in the auditory learning condition; non-musicians' pitch accuracy did not differ across learning conditions. Pianists and non-musicians engaged the superior temporal gyrus during auditory and motor (both minus cue-only) learning, and primary motor cortex (M1) during motor learning. Pianists showed greater M1 response during recall in the auditory learning condition compared to recall in the motor learning condition (both compared to recall during auditory-motor learning). Results suggest an expertise-dependent advantage for auditory over motor learning in auditory-motor skill acquisition. Skilled performers' sensory-to-motor connections may enable motor learning based solely on perception.

Artistic practice: Expressive goals provide structure for the perceptual and motor components of music performance

Sarah E. Allen[1], Carla D. Cash[2], Amy L. Simmons[3], Robert A. Duke[3]
[1]Southern Methodist University, USA
[2]Texas Tech University, USA
[3]The University of Texas at Austin, USA

The acquisition and refinement of the physical and perceptual skills involved in music performance demand both the automatization of fundamental motor behaviors and selective, focused attention to key aspects of proprioceptive and auditory feedback. The apportioning of conscious attention is central to effective practice in any domain, but seems especially challenging as musicians combine these components to instantiate expressive intention. In fact, the ways in which *thinking and attention* accommodate the various afferent and efferent signaling processes engaged during music making are not well understood. Even musicians who have achieved high levels of skill often approach individual components of music making independently during practice, without sufficient attention to the musical effects the physical skills of music making are intended to bring about, rendering their work ineffective in improving performance and overall artistry. In a series of experiments, we analyzed the practice of artist-level musicians and their students. Our results reveal that artist-level musicians focus attention on the expressive intentions of music making *during practice* in ways that serve to unify the physical components of music performance. We show in data from behavioral observations and interviews that the processes that underlie music learning are enhanced when musicians retain the integration of multiple aspects of playing and focus primarily on the accomplishment of expressive goals. Thus, the expressive aspects of music performance are not something "added later" once technical demands have been met. Instead, expression serves as the unifying element around which all of the physical and perceptual tasks cohere. In this session, we present recorded examples of effective music practice by artist-level musicians and highly accomplished students, and we discuss the approaches to music learning that facilitate the successful accomplishment of learning goals.

Attentional Control and Music Performance Anxiety: Underlying Causes of Reduced Performance Quality

Sarah E. Lade[1], Serena W. Tse[1], and Laurel J. Trainor[1]
[1]Department of Psychology, Neuroscience and Behaviour, McMaster University, Hamilton, ON, Canada

High-level music performance requires a high degree of attentional control. Attention is a mediating factor between performance quality and anxiety. Many musicians experience anxiety before a musical performance. Small amounts of anxiety are beneficial, allowing the musician to perform at his best. Some musicians, however, have an exaggerated stress response (Music Performance Anxiety, "MPA") and the quality of their performance suffers. The DSM-V categorizes MPA as a subtype of Social Anxiety (SA). Social phobics show lower attentional control following a socially stressful situation; however, no research exists on attentional control and MPA. Thus, the present experiment seeks to 1) determine whether MPA impacts attentional control and if so, 2) whether the resulting deficits are domain specific or domain general. Participants are screened with a standard MPA questionnaire (Perf-AIM). To induce stress, musicians (with and without MPA) perform in front of an audience. Participants then complete several attention tasks. The first three examine domain general processes. The Attention Network Task evaluates core attentional processes (alerting, orienting, executive function). The Wisconsin Card Sort Task evaluates positive attentional control, the ability to actively manipulate attention. An Emotional Stroop task evaluates negative attentional control, the ability to inhibit unwanted responses. We have also developed a second MPA/SA-specific Emotional Stroop task that examines domain specific processes. We predict that musicians with MPA will perform worse than non-MPA musicians on all attention tasks. We also predict that MPA musicians will perform worse on domain specific tasks than domain general tasks. It is unclear whether positive or negative attentional control will be implicated differently. This experiment is novel in investigating attentional control deficits as a potential underlying cause of degraded performance quality in MPA, in addition to comparing its effects to those of SA.

Effect of room acoustics on listeners' preference of string quartet performances recorded in virtual acoustics

Song Hui Chon[1], Sungyoung Kim[2], Doyuen Ko[3]
[1]School of Music, Ohio State University, Columbus, Ohio, USA
[2]Electrical, Computer & Telecommunications Engineering Technology, Rochester Institute of Technology, Rochester, New York, USA
[3]Audio Engineering Technology, Belmont University, Nashville, Tennessee, USA

Background
A musician's ability to present and perceive their own sound is greatly influenced by the acoustical properties of a given performance space. As various musical genres are performed in multi-purpose venues, today's advanced electro-acoustic technology has encouraged the development of virtual acoustic systems. Virtual acoustics offers flexibility and efficiency in adjusting the room acoustics of a concert venue to best accommodate musicians' unique needs of each performance.

Aims
This paper investigates listeners' preference of musical performances recorded in three different acoustic conditions, including two conditions with virtual enhancements. The goal of the study is to learn whether listeners, unaware of different acoustic conditions, prefer recordings in enhanced acoustic conditions, which performers reported favoring in our earlier study.

Method
Five professional string quartets were invited to record short excerpts in three different acoustic conditions. These excerpts were used for the listening test employing a forced-choice paired comparison paradigm. A group of amateur and professional musicians participated. Stimuli were presented over the headphones in a quiet laboratory, and listeners indicated which one of the two given excerpts they preferred, by how much, and why. 15 excerpts were tested twice for participant reliability verification.

Results & Conclusions
Data collection is ongoing. Based on our earlier studies showing that performers enjoyed the virtually enhanced acoustics that they were performing in, as well as that listeners could distinguish different acoustic conditions in recorded excerpts, we expect listeners to prefer performances recorded in virtually enhanced acoustic conditions rather than those in the natural condition. We will relate the listeners' preference ratings and the performers' preferences, then discuss the implication of acoustics on performance and its perception.

Computational analysis of expressive shape in music performance

Taehun Kim and Stefan Weinzierl

***Abstract*—**We present a computational method of expressive shape analysis in a music performance. Parameters in expressive performance are usually very fluctuating, but it is often possible to identify a curve-shaped prototype in it. Analyzing such a "shape" is useful to understand musical expression, but it is usually done by music experts. A computational analysis method is therefore significantly useful for developing performance-related computer systems. An expressive parameter fluctuation can be explained as a linear combination of shape and deviation. Assuming that the deviation follows a Gaussian distribution, we can compute the shape by means of Generalized Additive Model with Penalized Spline Regression. Since the shape is modelled with a sum of cubic splines, we can easily compute its derivatives, which provide detailed information about the shape. Visualization of analysis results shows the shapes estimated with our methods were fitted well to their original performances. Listening test results indicate the shape sounded expressive enough as its original. Computational analysis with our method was able to make same conclusion as a manual analysis made by music experts. In addition, we were able to identify two different interpretations from 4 professional performances of Chopin Prelude Op. 28 No. 7 by computational shape analysis.

***Keywords*—**expressive shape, interpretation study, music performance, performance analysis.

I. Introduction

EXPRESSIVE parameters in a music performance are usually very fluctuating. However, it is often possible to identify prototypes in curved shapes from those fluctuations. Such prototypes have been discussed in various music studies. B. Repp discusses commonality and diversity of music interpretation by means of shape analysis with parabolic functions [1]. P. Juslin *et al.* propose GERM model for musical emotion, and discuss human motion principles with curved shapes [2]. N. Todd argues that musical motion can be best described by parabolic functions [3]. Those studies show that the shape analysis in music performance is very useful to understand musical expression. However, they are usually done by music experts in a time-consuming manner, and they often require extra information from composition, such as measures and phrases. A computational method for expressive shape analysis in music performance is therefore significantly useful to develop computer systems for performed music.

Taehun Kim is with Audio Communication Group, Technische Universität Berlin, Berlin, Germany (corresponding author to provide e-mail: t.kim@mailbox.tu-berlin.de).

Stefan Weinzierl is with Audio Communication Group, Technische Universität Berlin, Berlin, Germany (e-mail: stefan.weinzierl@tu-berlin.de).

We present a computational method to analyze expressive shape from a music performance. An expressive parameter fluctuation can be explained as a linear combination of shape and deviation. Assuming that the deviation is following a Gaussian distribution, we can compute the shape by means of Generalized Additive Model (GAM) with Penalized Spline Regression. GAM allows us to define the shape as a sum of cubic splines, which is mathematically continuous and differentiable. This means that we can easily compute *extrema* and their locations in the shape, which represent musical tension and relaxation conveying performer's musical interpretation. In addition, the method we present does not require compositional features, and it is able to compute performances with any lengths. We visualize some results of computational shape analysis, and discuss result of listening test. In addition, we compare analysis results by our method and music experts, and show how to identify different interpretations from professional performances computationally.

II. Related Works

B. Repp shows that different parabolic curves can be found from different professional performances, but the model fitting is performed on selected excerpts like end of phrases [1]. A. Tobudic *et al.* propose a method with polynomial regression to compute phrasing from given performance, but it requires phrase information acquired by manual analysis [4]. S. Morita *et al.* propose a method with natural spline regression models [5]. The method requires so called *knots*, which defines the spline regression model's complexity, and they are manually given on every *n*-th note. P. Mavromatis proposes a non-parametric method based on *Parzen Window* approach [6]. He computes a non-parametric regression model with weighted sum of target parameter values in Gaussian Parzen window. The window size is estimated with Generalized Cross-Validation (GCV) [7], but the size is defined over the entire performance, so it can be understood as a filter with a fixed cutoff frequency. Most of those works require extra information from composition such as phrases, which makes it difficult to take account of interpretational diversity of music performance.

III. Method

A. Model of Expressive Shape

Assuming that a performer renders a musical piece with a prototype in a curved shape and spontaneous accents upon it, a performance can be represented with a linear combination of such a prototype and deviation. This can be mathematically expressed such as

$$y(t) = x(t) + r(t), \tag{1}$$

where $y(t)$, $x(t)$ and $r(t)$ denote original fluctuation of a performance parameter[1], prototype in a curved shape and its deviation, respectively. The prototype we will call expressive "shape", because it illustrates how performers shape musical expression in their interpretations. The deviation we will call expressive "deviation", because it is a small difference from the shape that makes musical accents. Computational analysis of the shape is then equivalent to a regression problem finding a model of $x(t)$, which fits best to the original parameter fluctuation, $y(t)$.

Solving that regression problem is not easy, because expressive shape is usually composed with complex curves, and we don't have much knowledge about its model. Tough, it seems plausible to assume that:

- a priori distribution of expressive deviation is a Gaussian one,
- expressive shape is explained with one of the spline regression family.

These assumptions allow us to compute expressive shape by finding an optimal spline regression model that keeps its residual following a Gaussian distribution, while the model best predicts the original parameter fluctuation. Fortunately, we can solve this regression problem by means of Generalized Additive Model with Penalized Spline Regression.

B. Computational Analysis with Generalized Additive Model

Generalized Additive Model (GAM) is a generalized linear model with a linear predictor involving a sum of smooth functions [8]. We redefine $y(t)$ in (1) by introducing a smooth function $f(\cdot)$ such as

$$y(t) = f(x_t) + \varepsilon_t, \tag{2}$$

where ε_t is an i.i.d. random variable following a Gaussian distribution, $\mathcal{N}(0, \sigma^2)$. According to the GAM theory, $f(x_t)$ can be defined with a sum of *basis functions* such as

$$f(x_t) = \sum_{j=1}^{q} b_j(\mathrm{t})\beta_j, \tag{3}$$

where $b_j(t)$ and β_j are the j-th *basis function* and its weight, respectively. The basis function can be any kind of mathematical functions, but we choose *cubic spline* as our basis function. This seems plausible, because we assumed that $x(t)$ in (1) is consist of complex curves, and it is known that musical motion can be well described with parabolic functions [3]. Then, we can rewrite (3) such as

[1] In case of piano performance, this would be tempo, dynamics or articulation

$$f(x_t) = \sum_{j=1}^{q} R(x, x_j^*)\beta_j, \tag{4}$$

where $R(x, x_j^*)$ denotes a *cubic spline* on j-th *knot*. The regression problem fitting $f(x_t)$ to $y(t)$ is then equivalent to determining the degree of smoothing. In case of natural spline regression models, we usually control the degree of smoothing with number of *knots* and their locations, which requires extra compositional features such as phrases and measures. However, the idea of Penalized Spline Regression provides an efficient solution for the problem without such extra information, by adding "wiggliness" penalty to the least square fitting objective. Rewriting (2) in a matrix form such as $\boldsymbol{y} = \boldsymbol{X}\beta + \varepsilon$, we can solve that regression problem by minimizing

$$\|\boldsymbol{y} - \boldsymbol{X}\beta\|^2 + \lambda \int_0^1 [f''(x)]^2\, dx, \tag{5}$$

where the integrated square of second derivative penalizes models that are too "wiggly". The smoothing parameter λ controls the trade off between model fit and smoothness. Therefore, we can find the best fit by calculating an optimal λ, where a sufficiently enough number of basis functions q is given. We set q with the number of quarter beats in given performances, that means the *knots* are set on every quarter beat, so that we have $q - 2$ number of cubic splines to define $f(x_t)$.

The optimal λ can be found by searching the model fit that best predicts the original parameter fluctuation. This can be achieved by cross validation method. Having performance parameter value on every n-th note, we can define *ordinary cross validation* score $\mathcal{V}_o$ such as

$$\mathcal{V}_o = \frac{1}{N}\sum_{n=1}^{N}(\hat{f}^{[-n]}(x_n) - y(n))^2, \tag{6}$$

where $\hat{f}^{[-n]}(\cdot)$ represents a regression model fitted to all of original data points except the datum on n. $\mathcal{V}_o$ can be interpreted as an average prediction error of models created by *leave-one-out* method on every parameter value $y(n)$. We choose then the optimal λ, which minimises $\mathcal{V}_o$. This can be efficiently computed by GCV with Penalized Iteratively Re-weighted Least Squares algorithm known as P-IRLS [9]. In this way, we can find the optimal fit, $f(x_t)$, which best predicts the original data points $y(t)$, while keeping ε_t follows a Gaussian distribution, $\mathcal{N}(0, \sigma^2)$.

The model of expressive shape, $f(x_t)$, is a sum of cubic splines. A cubic spline is a mathematically continuous and differentiable function, so that its sum $f(x_t)$ is also continuous and differentiable. Therefore, we can easily compute its derivatives, which provide detailed information about the shape. For example, zero crosses of $f'(x_t)$ indicate where musical tension and relaxation happen.

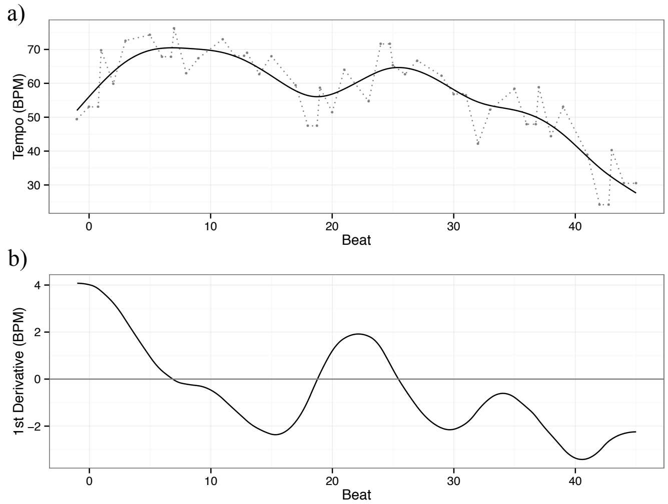


Fig. 1 Tempo shape analysis of F. Chopin Prelude Op. 28 No. 7 by Vladimir Ashkenazy. a) expressive shape (solid line) fitted to the original fluctuation (dotted line). b) the first derivative of the shape.

IV. RESULT

A. Expressive Shape Analysis

Fig. 1 shows computational tempo shape analysis result of Chopin Prelude Op. 28 No. 7 performed by V. Ashkenazy[2] with our method. Fig. 1a shows the expressive shape was in a form of a smoothed curve, and fitted well to the original tempo fluctuation. The shape informs there were two accelerations and decelerations during the performance. The first deceleration in the middle divides the performance in two sections, and the other one on tail of the performance implements final *ritardando*. Fig. 2b shows the first derivative of the computed shape. The zero crosses from positive to negative values indicate tensed positions, and ones from negative to positive values indicate relaxed positions. The derivative shows one relaxed position in the middle of the performance, which indicates a boundary of those 2 sections in the interpretation.

Fig. 2 depicts another shape analysis results of 3 different performances of compositions from various musical epochs. The result shows that the expressive shapes computed with our method were also fitted well to the original tempo fluctuations in all of those performances.

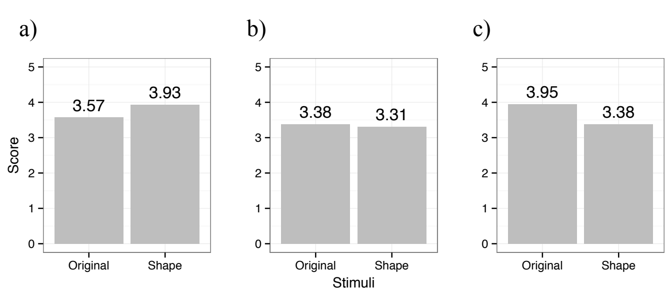


Fig. 3 Listening test results of a) *cho-mzk019-czern-y*, b) *moz-snt279-2-pires-y* and c) *rav-sonatn1-rogep-y*.

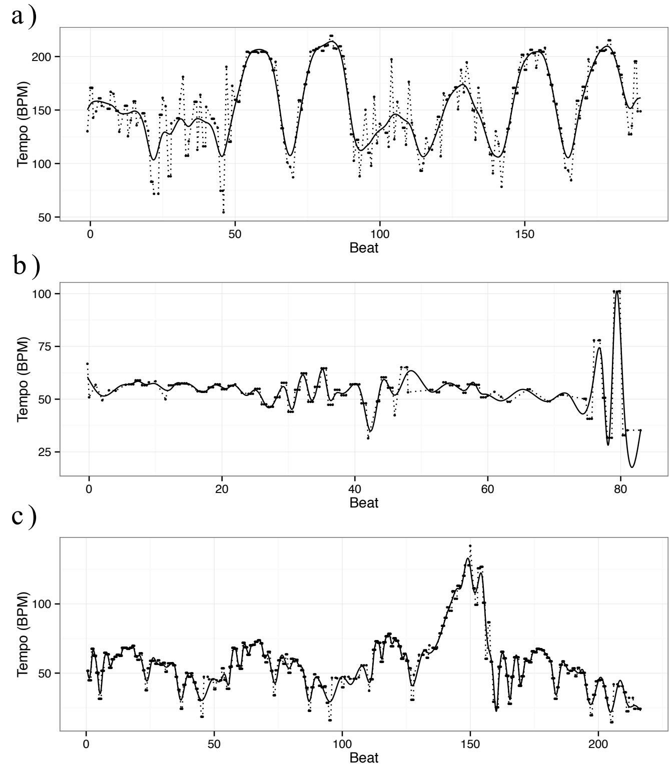


Fig. 2 Tempo shape analysis of diverse performances from CrestMuse PEDB. a) F. Chopin, Mazurka Op. 30 No. 2 performed by H. Czerny-Stefanska (*cho-mzk019-czern-y*), b) W. A. Mozart, Piano Sonata KV279 II. Andante performed by M. Pires (*moz-snt279-2-pires-y*), c) M. Ravel, Sonatine No. 1 "Modéré" performed by R. Peltzmann (*rav-sonatn1-rogep-y*).

B. Listening Test

We conducted listening tests to understand expressivity of the shapes computed by our method. We had 42 subjects[3], who listened to the 3 stimuli such as original performance, *shape*-only and *deviation*-only[4] renditions, which were derived from 3 performances depicted in Fig. 2. The subjects firstly listened to the original performance, and then listened to the 2 other stimuli. The type of the stimulus presented to the listeners was blind. After listening they evaluated expressivity of those 3 stimuli with 5 level *Likert Scale*, which represents a scale from *not expressive at all* to *very expressive*.

Fig. 3 shows the mean scores of original performance and *shape*-only rendition derived from those 3 performances. The mean score differences of Chopin Mazurka (Fig. 3a) and Mozart Piano Sonata (Fig. 3b) were not significant by ANOVA and TukeyHSD post-hoc tests ($p = 0.20$ and $p = 0.94$, respectively). This indicates the subjects perceived *shape*-only rendition expressive enough as its original performance. Although the score difference was not significant, some subjects found that the *shape*-only rendition conveyed performer's artistic intention better than the original performance of Chopin Mazurka. This result addresses a

[2] *cho-pld007-ashke-b* from CrestMuse PEDB. More information about the database can be found in http://www.crestmuse.jp/pedb/

[3] 5 professionals, 26 hobbyists and 11 others.

[4] *Deviation* is not discussed here, because it is out of scope of this paper.

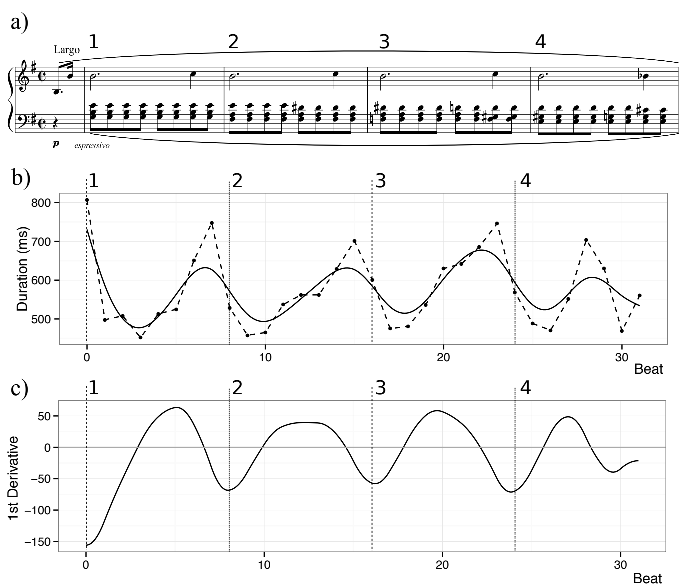


Fig. 4 a) Score of the first 4 bars in Chopin Prelude Op. 28 No. 4. b) note durations in the performance of Martha Argerich, which were measured by Senn *et al*. (dotted line), and shape analysis result computed by our method (solid line). c) the first derivative of the expressive shape.

hypothesis that expressive shape would be one of the most important components in performer-listener communication, which requires further researches though.

In case of Ravel Sonatine, the score of *shape*-only rendition was significantly lower than its original performance ($p = 0.01$). However, the score of *shape*-only rendition was higher than 3, which means the subjects found that it sounded expressive with certain degree. This result does not indicate that our method failed to analyse, because the roles of expressive shape and deviation for building entire expression could be varying for different epochs and styles. Still, it addresses limitations of our method for the performances of music pieces in the recent music history such as in the first half of the 20th. century.

C. Comparison with Analysis by Experts

Senn *et al.* published a tempo analysis of Chopin Prelude Op. 28 No. 4 performed by Martha Argerich [10]. Fig. 4 shows performed duration of each eighth note in the first 4 bars[5]. They found significantly prolonged eighth notes in the end of each bar, and call it "bar line *ritardando*". However, they found an exceptional *ritardando* in the middle of the fourth bar. They argue that it is intended by the performer and seems plausible, because the last note of the fourth bar functions as a preparation for the start of the next melodic segment.

Fig. 4b shows expressive shape computed by our method, which illustrates rough duration changes in a curved shape. The shape illustrates well how the duration changes over time. In bar 1-3 there are peaks in the end of each measure,

[5] We reproduced the measurement from Senn *et al.* as a line graph. The raw data you can find from [10].

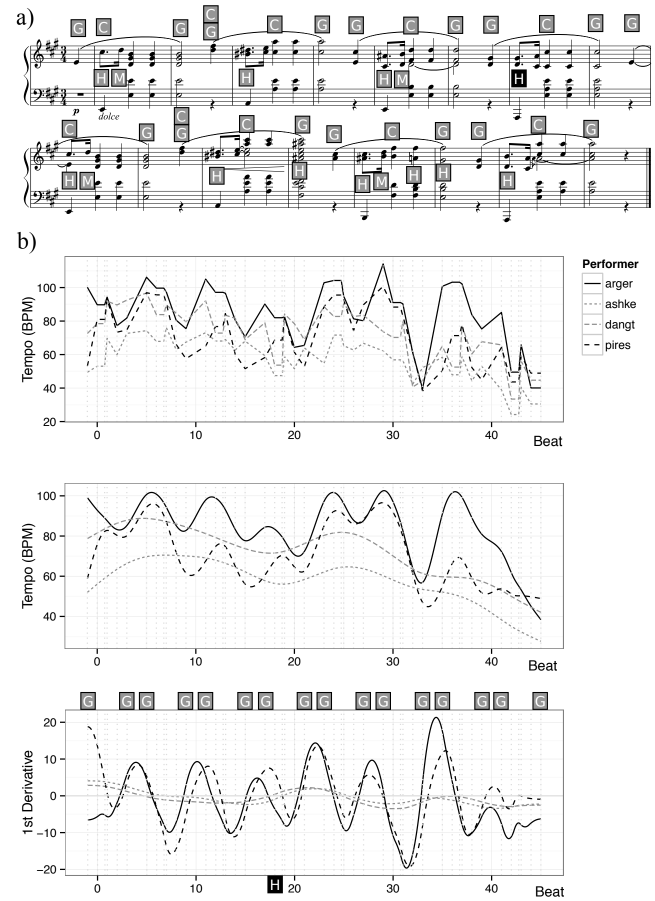


Fig. 5 Four performances of Chopin Prelude Op. 28 No. 7. a) *accents* analysis of the composition by R. Parncutt published in [11]. b) original tempo fluctuations, expressive shapes and their first derivatives.

but in the fourth bar there is a peak in the middle of the measure. This confirms what Senn *et al.* found in their analysis. Fig. 4c shows the first derivative of the shape. The zero-crosses from positive to negative values indicate those peaks, where the tempo is mostly slowed down. Such zero-crosses are observable near the last eighth note in the bar 1-3, and in the middle of the fourth bar. This indicate that we were able to draw the same conclusion by the computational analysis as Senn *et al.* did.

D. Identification of Different Interpretations

In order to learn potentials of the computational expressive shape analysis, we tried to identify different interpretations from professional piano performances with our method. We selected 4 performances of Chopin Prelude Op. 28 No. 4 by M. Argerich (*arger*), V. Ashkenazy (*ashke*), Dang Thai Son (*dangt*) and M. Pires (*pires*), and performed computational shape analysis on them.

Fig. 5 shows tempo changes of those performances. Just having a look at the original fluctuations, it is not easy to find interpretation styles. However, the shape analysis shows clear tempo accelerations and decelerations, and reveals two different groups: one with frequent periodical relaxations, and the other one with a single relaxation in the middle of the performance (Fig. 5b). The first derivatives can be easily obtained from those shapes, and their zero crosses indicate where relaxations happen.

R. Parncutt *et al.* analyzed musical accents of the composition regarding melodic contour (C), grouping (G), meter (M) and harmony (H), which is useful to understand the interpretations identified with computational shape analysis [11]. Considering the composition and shape analysis all together, we were able to identify two different interpretations: M. Argerich and M. Pires emphasize repeating rhythmic motif by tempo slowed down on grouping accents (G), and V. Ashkenazy and Dang Thai Son divide the piece into two sections, whose boundary is emphasized by tempo slowed down on the harmonic accent (H) in the middle of the piece.

V. SUMMARY

We presented a computational method of expressive shape analysis in music performance, and showed several results indicating plausibility of the computational analysis. As the next step, we will perform a large scale computational analysis on some of published performance databases, and discover its potentials for performance-related computer systems. Furthermore, we will research about integration of the computational analysis method to artificial music performance systems for learning and generating expressive shapes.

ACKNOWLEDGMENT

This research is partially supported by Samsung Scholarship Foundation.

REFERENCES

[1] B. Repp, “Diversity and commonality in music performance: An analysis of timing microstructure in Schumann’s Träumerei,” in *Journal of the Acoustical Society of America*, vol. 92(5), 1992.

[2] P. Juslin and R. Timmers, “Expression and communication of emotion in music performance,” in *Handbook of music and emotion: theory, research, applications*. London: Oxford University Press, 2010.

[3] N.P. Tood, “Kinematics of musical expression,” in *Journal of the Acoustical Society of America*, vol. 97(3), 1995.

[4] A. Tobudic and G. Widmer, “Playing Mozart phrase by phrase,” in *Proc. International Conference on Case-Based Reasoning*, 2003.

[5] S. Morita, N. Emura, M. Miura, S. Akinaga and M. Yonagida, “Evaluation of a scale performance on the piano using spline and regression models,” in *Proc. International Symposium on Performance Science*, 2009.

[6] P. Mavromatis, “A Multi-tiered approach for analyzing expressive timing in music performance,” in *Mathematics and Computation in Music*, Springer Verlag, 2009.

[7] G. Golub, M. Heath and G. Wahba, “Generalized Cross-Validation as a method for choosing a good ridge parameter,” in *Technometrics*, vol. 21(2), 1979.

[8] T. Hastie and R. Tibschirani, “Generalized Additive Models,” in *Statistical Science*, vol. 1(3), 1986.

[9] C. Gu, “Cross-validating non-gaussian data,” in *Journal of Computational and Graphical Statistics*, vol. 1(2), 1992.

[10] O. Senn, L. Kilchenmann and M.-A. Camp “Expressive timing: Martha Argerich plays Chopin's Prelude op. 28/4 in E minor,” in *Proc. International Symposium on Performance Science*, 2009.

[11] J. MacRitchie, “Elucidating musical structure through empirical measurement of performance parameters,” Ph.D. dissertation, University of Glasgow, 2011.

Solo vs. duet in different virtual rooms: On the consistency of singing quality across conditions

Timo Fischinger[1], Gunter Kreutz[2] , Pauline Larrouy-Maestri[3]
[1]Music Department, Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, Germany
[2]Institute for Music, Carl von Ossietzky University of Oldenburg, Oldenburg, Germany
[3]Neuroscience Department, Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, Germany

Previous research on vocal pitch accuracy revealed insights into the fundamentals of singing. However, most of the research on singing focused on the analysis of single voices, whereas few attempts have been made to tackle the challenge of analyzing *multitrack* recordings of singing ensembles. In addition, singers have to adjust their way of singing with respect to a given venue's acoustical environment (e.g., small room vs. a comparatively large space like a church). If it is common that musical performances are greatly influenced by room acoustics, studies on the effects of room acoustical features during ensemble singing are rare. In order to investigate singing performances across various conditions, we manipulated the singing condition (unison, canon, solo) as well as the acoustical feedback by applying diverging virtual rooms. Three duets with female singers (N = 6) were asked to sing three different melodies using headset microphones to record each singer separately. Recordings took place in the communication acoustic simulator (CAS) at the House of Hearing (Oldenburg, Germany) to be able to provide different simulated acoustical spaces (i.e., cathedral, classroom, and dry condition) to the singers. Objective measures were performed on each recording and confirmed that the singers sang the melodies with high precision (small pitch interval deviations) hardly affected by singing conditions or by the type of acoustical feedback. However, the singers tended to drift (larger deviations of the tonal center) when singing in canon compared to solo and unison singing. Overall, the analysis of the pitch accuracy showed a general effect of condition (i.e., unison, canon, solo), but no general effect of acoustical feedback and no interaction between the two variables under study.

The relaxing effect of tempo on music-aroused heart rate

Edith Van Dyck[1], Joren Six[1] and Marc Leman[1]
[1] IPEM, Department of Arts, Music and Theater Sciences, Ghent University, Ghent, Belgium

Background
Music is frequently used as a means to relax, while its arousing effects are often employed in sports and exercise contexts. Previous work shows that music tempo is one of the most significant determinants of music-related arousal and relaxation effects. However, in the literature on human heart rate, music tempo itself has not yet been studied rigorously on its own.

Aims
The aim was to investigate the link between music tempo and heart rate during passive music listening, adopting an experimental design with tight control over the variables.

Method
Heart rate was measured in silence after which music was provided in a tempo corresponding to the assessed heart rate. Finally, the same stimulus was presented again but music tempo was decreased, increased, or kept stable.

Results
The experiment with heart rate measurements of 32 participants revealed that substantial decreases in music tempo significantly reduced participants' heart rates, while heart rate did not respond to less considerable drops or increases. It was also shown that heart rate significantly increased in the music condition compared to the silent one. The person's gender or music preference did not seem to be of significant importance.

Conclusions
Generally, it is believed that music can induce measurable and reproducible effects on human heart rate, leading to a condition of arousal proportional to the tempo of the music. However, our findings revealed that only substantial decreases in music tempo could account for heart rate reductions, while no link between tempo increase and heart rate was uncovered. As music listening showed to increase heart rate compared to silence, it is suggested that possible effects of tempo increases are regulated by the arousal effect of music itself. These results are a major contribution to the way in which music is used in everyday activities and are valuable in therapeutic and exercise contexts.

Entrainment-based Music Interventions for Surgery: Harnessing the Iso Principle

Courtney Williams[1]

[1]Queensland Conservatorium, Griffith University, South Brisbane, Queensland, Australia; email: courtney.williams3@griffithuni.edu.au

Background

Music has been shown to positively affect a patient's mental and physiological state in perioperative settings. This effect is often explained by entrainment – the synchronization of the patient's bodily rhythms to rhythms presented in the music. While there is evidence to support the effectiveness of musical entrainment, there exists a need for a more specialized investigation into this field, which moves beyond the basic discussions of tempo and looks at musical selections at both an individual level, and as a whole.

Aims

This research examines the literature surrounding the use of entrainment in music-based surgical environments, creating a method for the selection and composition of music for this purpose.

Main Contribution

It is argued that music can not only entrain the rhythms of the body to states more conducive to healing, but that the mental state of the listener can also be entrained by the moods presented in the music. As such, it is advocated that these two types of entrainment should be referred to using a more specialized terminology – 'physiological entrainment' and 'emotional entrainment' respectively. It also argued that the iso principle – an established music-based technique for the targeted shifting of emotional states – can be applied not only to emotional entrainment interventions, but can also be translated to physiological entrainment interventions.

Implications

This research will aid future musical entrainment interventions by providing a clear method for the selection or composition of the most effective and appropriate music for this purpose. Standardizing music selection in this way will contribute to the improvement of the reliability of studies, and thus will aid in enhancing future research in this field.

The impacts of music-supported rehabilitation on auditory evoked magnetic fields in single stroke cases

Claudia Freigang[1], Bernhard Ross[1,2], and Takako Fujioka[3]
[1]Rotman Research Institute, Baycrest Hospital, Toronto, ON, Canada
[2]Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada
[3]Department of Music and Stanford Neuroscience Institute, Stanford University, California, USA

Stroke is the main cause for adult disability, leading to the loss of motor skills. A recently developed therapy, music-supported rehabilitation (MSR), utilizes goals of music-making activities with percussion and keyboard instruments to improve motor abilities in stroke patients, and the respective results have been promising. Multiple studies have shown that musical training leads to neuroplastic changes in auditory and motor domains. The effect of MSR on the auditory system is not clear yet. The present study investigated the effect of MSR on auditory cortical processing in chronic stroke patients by means of magnetoencephalography (MEG) recordings by examining mismatch negativity (MMN) of auditory evoked fields (AEF). Lesion size and location varies largely across stroke patients, thus we evaluated the outcome of MSR and a conventional therapy (CT) for single cases. 29 patients participated in MSR or CT for 15 or 30 times of one-on-one 1-hour session with a therapist. We administered our test battery at three/four time points: before training, after 15h or 30h of training, and 3 months after the last training session. The test battery consisted of MEG recordings and behavioural tests. MEG recordings entailed: (a) Presentation of single complex tones to obtain standard AEF responses; and (b) passive and active oddball tasks to extract MMN responses. Previous studies found improved MMN responses after musical training. We hypothesized that MMN will change in the patients receiving MSR, but not in those receiving CT. Behavioural tests assessed auditory, motor and cognitive performance. The different lesion locations revealed a large variability on AEF responses to a single complex tone. Thus, we discuss the results of MSR-dependent changes in AEFs and motor abilities based on lesion side. Motor rehabilitation will be discussed for single stroke cases by comparing the outcomes for MSR and CT. Limitations in evaluating AEFs in stroke patients will also be discussed.

Neuroimaging and therapeutic use of psychedelic drugs and music

Mendel Kaelen[1], Romy Lorenz[2], Anna Simmonds[2], Leor Roseman[1,2], Amanda Feilding[1], Robert Leech[2], David Nutt[1] and Robin Carhart-Harris[1]
[1]Centre for Neuropsychopharmacology, Imperial College, London, United Kingdom
[2]The Computational, Cognitive and Clinical Neuroimaging Laboratory, Imperial College London, UK

Background
Therapeutic potentials of psychedelic drugs such as LSD and psilocybin are being re-examined today with promising results. During psychedelic therapy patients typically listen to music, which is considered an important element to facilitate therapeutic experiences, often characterized by profound emotion. Underlying mechanisms have however never been studied.

Aims
Studies aimed 1) to understand how music and psychedelics work together in the brain, and 2) how their combination can be therapeutic.

Method
1) The interaction between LSD and music was studied via a placebo-controlled functional magnetic resonance imaging (fMRI) study. 20 healthy volunteers received LSD (75mcg) and, on a separate occasion, placebo, before scanning under resting conditions with and without music. Changes in functional and effective connectivity were assessed, alongside measures of subjective experience. In addition, it was studied how LSD alters how the brain processes acoustic features (timbre, tonality and rhythm).
2) In a clinical trial, psilocybin was used to treat 20 patients with major depression. Semi-structured interviews assessed the therapeutic role of music, and fMRI scanning before and after treatment assessed enduring changes in music-evoked emotion.

Results
1) Compared to placebo, changes in functional and effective connectivity were observed during LSD in response to music, including altered coupling of the medial prefrontal cortex (mPFC) and parahippocampus. These changes correlated with subjective effects, including increased mental imagery. 2) Music influenced the therapeutic experience in several ways, including enhancing emotion. Altered functional connectivity in response to music after treatment was associated with reductions in depressive symptoms.

Conclusions
These findings suggest plausible mechanisms by which psychedelics and music interact to enhance subjective experiences that may be useful in a therapeutic context.

Benefits of a Voice Therapy and Song Singing Group for People with ILD

Merrill Tanner,[*1] Meena Kalluri,[#2] Janice Richman-Eisenstat[*3]

[*]*Glenrose Rehabilitation Hospital, Edmonton, Alberta, Canada*

[#]*Dept. of Medicine, University of Alberta, Edmonton, Alberta, Canada*

[*]*Dept. of Medicine, University of Alberta, Edmonton, Alberta, Canada*

[1]merrill.tanner@ualberta.ca, [2]kalluri@ualberta.ca, [3]richmane@ualberta.ca

ABSTRACT

Interstitial lung diseases (ILD) are characterized by fibrosis and impaired gas exchange in the lungs. Shortness of breath and cough are the main symptoms. These respiratory diseases are complex and can be difficult to treat. A singing program led by a Speech-Language Pathologist/Singer was provided to 4 people with ILD. 8 sessions of voice therapy exercises and song singing with piano accompaniment took place over a 4 week period. The goal of the study was to find out whether singing and voice therapy would be beneficial to people with ILD. Pre and post measures were taken. Respiratory measures included assessment of forced vital capacity, and a rating on the Dyspnea Medical Research Council breathlessness 1-5 scale and cough rating scale (mild, moderate, severe). Acoustic measures included average pitch in reading, maximum phonation time (MPT), loudness range and pitch range. Unsolicited comments given during post testing were also recorded. Lung function did not change in two patients and improved in one patient. The Dyspnea MRC breathlessness scale improved in 2 patients and remained the same in another 2 cases. In one of the 2 patients that had a cough, a mild cough disappeared. Pre-post acoustic measures showed that average pitch in reading rose and that MPT improved for all participants. Loudness range, pitch range and perceptual ratings improved in ¾ of participants. Comments of participants during post testing revealed that all of the participants enjoyed the group and felt that their voices had improved. 3 of the 4 participants felt that their breathing had improved, and 2 reported less coughing. 3 mentioned that they would like the group to continue as a part of the pulmonary rehabilitation program. One participant returned to singing in a choir after the end of the session. Respiratory, vocal and mood improvements were evident in this small group of four participants with ILD. Voice therapy with singing appears to improve the lives of people with ILD and should be studied further.

I. BACKGROUND

Idiopathic pulmonary fibrosis (IPF) is a chronic, progressive, irreversible, and fatal fibrotic interstitial lung disease (ILD). It is a devastating condition that afflicts people with relentlessly progressive shortness of breath [1]. Despite 30 years of intensive research, there is no cure and mortality rate remains high [2]. Review of the literature on IPF related quality of life (QOL) reveals several unique as well as over lapping problems shared with other lung diseases. The hallmark symptom of IPF is exertional dyspnea, which is quite disabling in patients with IPF. It measurably worsens over time and appears to be one driver of reduced QOL.

In addition to dyspnea, IPF patients also suffer from sadness, worry, anxiety, fear, isolation and a sense of lack of control over their condition [3]. Patients commonly express hopelessness and despondency when they learn that there are no effective therapies for disabling symptoms. Physical limitations and oxygen dependency impose an overpowering sense of social and emotional isolation that seems to have not been captured by any QOL surveys done to date. Several QOL related studies have noted that the domains of mental health also show impairment [4]. It is known that depression and anxiety can have a negative impact on lung function, can worsen quality of life and also lead to higher healthcare costs [5].

The current therapies may slow down progression of disease but have no impact on symptoms or QOL. Therefore, the goal of care should be to help IPF patients experience less dyspnea, less anxiety, less depression, and to have a better quality of life. Pulmonary rehabilitation (PR) programs have been reported to improve quality of life in IPF [6]. They have also been associated with improvements in mood, depression and QOL in healthy volunteers and patients with COPD [8-9]. However, delivering effective psychological, emotional and social support still remains a challenge in IPF. The use of music therapy in the form of singing teaching can be an effective tool to achieve these ends. Voice training and singing has been shown to improve QOL and lung function.

Trials with COPD and asthma patients have shown that involvement in a choir group not only encourages social bonding and alleviates depression, anxiety, fear and isolation, but also gives a sense of purpose in life. A study of 43 COPD patients showed improvement in QOL, lung function and dyspnea [11]. In another study of 7 subjects with emphysema who participated in twice weekly singing classes for six weeks improvement in chest wall mechanics was observed [12]. A pilot study evaluated quality of life (QOL) and lung function before and after three months of choral singing in cancer survivors and their care givers and found improvements in QOL and depression. Structured interviews revealed that the choir provided focus; the participants felt uplifted and had greater confidence and self-esteem [13]. Another study of a group of Alzheimer's disease patients and their caregivers showed the efficacy of receptive music therapy in alleviating anxiety and depression [14].

II. RATIONALE

IPF is a chronic, progressive, fibrotic interstitial lung disease with mortality rates approaching those found in lung malignancy. IPF patients suffer from depression, anxiety, fear and social isolation, in addition to the progressive dyspnea. Despite intense research there is no curative therapy. There are several studies related to asthma, COPD and lung cancer that showed that participation in choir groups has effects beyond lung function. It promotes social and spiritual wellbeing and thereby improves quality of life. This has never been studied in relation to interstitial lung diseases like idiopathic pulmonary fibrosis (IPF).

III. STUDY OBJECTIVE

To conduct a pilot study that would assess the impact of participation in a choir group by patients with ILD (including IPF) on their quality of life, lung function and voice.

IV. METHODS

A small study with 4 people with ILD was conducted in August of 2015. Subjects were recruited from the Multidisciplinary ILD program by Dr. Kalluri. Dr. Richman-Eisenstat assessed them and determined their exercise prescription. Patients were offered participation in a singing group. Three of the four patients involved attended the pulmonary rehabilitation program as well. The program consisted of four weeks of twice weekly singing sessions led by a Speech-Language Pathologist/Singer and accompanied by piano. The sessions were initially 1/2 hour long but were expanded to one hour due to patient requests for longer sessions. Sessions involved voice exercises, including semi-occluded vocal tract exercises, open throat exercises, vocal hygiene, breathing techniques and song singing. Home practice was encouraged. Songs included common folk and older popular repertoire with positive themes. Attendance was variable as these patients were living in the community and trying to deal with work and the limitations of their disease.

A. Evaluation

Pulmonary measures were taken in clinic by the respiratory therapist before and after the treatment period. Voice measures were taken by the Speech Language Pathologist leading the group before and after the treatment period.

B. Results

1) Pulmonary Results

MRC dyspnea scale, CGI-I (clinical global impression of improvement) and objective measures of lung function such as pulmonary function test (FVC, DL) and 6 minute walk distance at the two time points were measured. €Two participants had improvements in functional vital capacity and reduced breathlessness. One participant had improved cough and another had an improved clinical global impression of improvement.

Name	FVC		Dyspnea MRC 1-5 scale		Cough mild- mod- severe		CGI-I
	Pre	Post	Pre	Post	Pre	Post	
1	2.73 104%	2.75 105%	4	3	None	none	4 (no change in overall clinical status)
2	1.87 43%	1.66 38%	4	4	Mild	None	4 (no change in overall clinical status)
3	2.40 79%	2.63 87%	3	3	Mild	Mild	4 (no change in overall clinical status)
4	2.14 64%	2.11 63%	3	2	None	None	2 (much improved compared to previous)

2) *Voice Results*

Average pitch during reading and maximum phonation time improved in all four cases. Loudness range, pitch range, and perceptual ratings by the therapist improved in three out of four cases. The initial pitch range reading of one patient may have been inaccurate due to the perturbation and shimmer in his voice. The equipment may have picked up high frequencies that were not a part of the perceived vocal range. Post testing for one patient occurred in her work environment where she was not able to be as loud as possible. In the perceptual measures the one patient who did not improve remained at the same as in the initial assessment. One person of the three that completed the forms, improved on both vocal QOL measures. The other two appeared to increase their scores (poorer vocal QOL). This may have occurred since their self awareness of problems increased as a result of the vocal hygiene education.

3) *Interview results*

All participants mentioned that they enjoyed the group and would like to continue. One person wanted it to increase to three times per week. Two people felt that their coughing had decreased. Another person felt that her throat was more relaxed. One participant said he would highly recommend the group to anybody and that his improved voice now allows him to sing along with music in the car. Two people felt that the quality of their voices was better in speaking and singing. Two others reported that their higher range had improved. One woman said she felt more confident, while another said that the group had improved her mood. She thought that meeting with others who had the same problem and having fun both contributed to her better mood. As a result of being in the group she returned to the choir that she attended before falling ill with ILD and found that she could participate if she took more frequent breaths that previously. The same woman felt that she now had improved breathing and was able to use the vocal techniques learned to recover more quickly when walking too fast. Another participant reported that she had learned methods to control voice and breathing.

C. **Discussion**

Despite the small size of group (n=4) and the bias in the voice results (the treating therapist also performed voice testing), the study appears to show that all participants experienced some benefit.

D. **References**

1. American Thoracic Society. IPF: diagnosis and treatment. International consensus statement. *Am J Resp Crit Care Med* 2000;161:646.
2. AJRCCM 1998:157:199-203-Flaherty.
3. Pulmonary rehabilitation in idiopathic pulmonary fibrosis: a call for continued investigation. *Resp medicine* 2008;102;12:1675.
4. Evaluation of the short-form 36-item questionnaire to measure health-related quality of life in patients with idiopathic pulmonary fibrosis. *Chest* 2000;117(6):1627.

5. The prevalence and pulmonary consequences of anxiety and depressive disorders in patients with asthma. *Coll Antropol.* 2012 Jun; 36(2):473-81.
6. Effectiveness of pulmonary rehabilitation in restrictive lung disease. *J Cardiopulm rehab* 2006; 26(4):237-243.
7. 'Didgeridoo playing and singing to support asthma management in Aboriginal Australians', *Journal of Rural Health,* vol. 26, no. 1, pp. 100-4.
8. The perceived benefits of singing: findings from preliminary surveys of a university college choral society. *J R SOC PROmotion Health* 2001, 121, 4, 248-256.
9. Effects of singing classes on pulmonary function and quality of life of COPD patients*. Intl j of copd* 2009:4;1-8
10. Karaoke for quads: A new application of an old recreation with potential therapeutic benefits for people with disabilities. *Disability and Rehabilitation*, 25, 297-301.
11. Singing teaching as a therapy for chronic respiratory disease - a randomised controlled trial and qualitative evaluation. *BMC pulm med* 2010; 10:41.
12. The singer's breath: implications for treatment of persons with emphysema. *J Music Ther.* 2005 Spring; 42(1):20-48.
13. A pilot investigation of quality of life and lung function following choral singing in cancer survivors and their carers. *Ecancer* 2012;6:261.
14. Impact of music therapy on anxiety and depression for patients with Alzheimer's disease and on the burden felt by the main caregiver (feasibility study). *Encephale* 2009, 35, 1, 57-65, France.

V. CONCLUSION

In the absence of pharmacological therapy and ILD tailored pulmonary rehabilitation, patients with IPF continue to suffer from breathlessness. Based on the beneficial initial findings of this pilot study there is reason to continue research on singing programs for Interstitial Lung Disease.

IV. ACKNOWLEDGMENT

Funding for this study was obtained from the Canadian Pulmonary Fibrosis Foundation (CPFF).

The Use of Therapeutic Singing Program to Enhance Vocal Quality and Alleviate Depression of Parkinson's Disease: Case series

Eun Young HAN[1,2], Ji Young YUN[2] , Kyoung-Gyu CHOI[2] and Hyun Ju CHONG[1]

[1] Department of Music Therapy, Ewha Womans University Graduate School, Seoul, Korea

[2] Department of Neurology, Ewha Womans University School of Medicine, Ewha Womans University Mokdong Hospital, Seoul, Korea

The symptoms of Parkinson's disease (PD) patients are not only movement disorder but also voice and affective disorder so that they frequently suffer emotional depression as well as communication problems. This depression causes reduced quality of life and delayed therapeutic responses. Music acts as a stimulus to elicit motor and emotional responses and has been applied to various neurologic disorders. Therapeutic singing activates neural mechanisms for speech using the respiratory muscles and the articulation. Singing and speaking are similar in that they involve cortical regions and engage bilateral activities of the brain. Moreover, the melody and lyrics of singing frequently induce personal memories related to both physical and emotional states. The purpose of this study is to examine the Therapeutic Singing Program (TSP) to enhance vocal quality and alleviate depression of PD. We studied 6 female and 3 male PD patients aged 53 to 78. They were selected by the neurologist with clinical criteria. Each patient received an intensive music therapy session for 50 minutes in a total of 6 sessions for 2 weeks. The therapeutic singing program consisted of patterned vocalizing, vocal improvisations and song compositions for vocal exercise. In the pre/post tests on the 1st and 8th days, we measured the Maximum Phonation Time (MPT) via the Praat test, the Voice Handicap Index (VHI), the Voice-Related Quality of Life (V-RQOL), and the Geriatric Depression Scale (GDS). We conducted the follow-up tests six months later. The vocal quality of the patients, including acoustic and subjective vocal evaluations, was improved and the depression symptoms of the patients were alleviated. We also found that both the acoustic vocal in the MPT and the V-RQOL were increased while the VHI was decreased. The GDS was found remarkably lowered. The follow-up tests indicated positive effects of the Therapeutic Singing Program on keeping the vocal quality and alleviating the depression. This holistic music therapy study for Parkinson's disease has confirmed the effectiveness of the Therapeutic Singing Program (TSP) to enhance vocal quality and alleviate depression of PD using vocalizing, vocal improvisation and song compositions based on phonetics, neuropsychology and music psychotherapy.

The Therapeutic Function of Music for Musical Contour Regulation Facilitation: A model to facilitate emotion regulation development in preschool-aged children

Kimberly Sena Moore[1], Deanna Hanson-Abromeit[2]
[1]Department of Music Education and Music Therapy, Frost School of Music, University of Miami, Coral Gables, Florida, USA
[1]Division of Music Education and Music Therapy, School of Music, University of Kansas, Lawrence, Kansas, USA

Emotion regulation (ER) is an umbrella term to describe interactive, goal-dependent explicit and implicit processes that are intended to help an individual manage and shift an emotional experience. The window for appropriate ER development occurs during infancy, toddlerhood, and preschool. Atypical ER development is considered a risk factor for mental health problems and has been implicated as a primary mechanism underlying childhood pathologies. Currently available treatments are primarily verbal- and behavioral-based. Furthermore, there is a disconnect between the timing of the treatment and the need to practice how to handle emotionally charged situations in the moment, as well as a lack of a caregiver-child interactive component. There is behavioral and neural support for using music as a therapeutic mechanism, thus the incorporation of intentional music experiences may be one way to address these limitations. The purpose of this research is to provide a theoretical rationale for the Therapeutic Function of Music (TFM) to support Musical Contour Regulation Facilitation (MCRF), a music-based intervention for emotion regulation development in preschoolers. The intention of the MCRF is to use the contour and temporal structure of a music therapy session to mirror the changing flow of the caregiver-child interaction through the alternation of high arousal and low arousal music experiences. A review and synthesis of the music theory, music neuroscience, and music development literature was conducted to inform the TFM. Results from the analysis provided preliminary guidelines to outline the structural characteristics of the music stimulus to create musically facilitated high and low arousal experiences. Developmentally appropriate music for preschool-aged children should include rhythmic and melodic repetition, consonant harmonies, binary rhythms, and an easy-to-follow, step-wise melodic contour that falls within an octave range. Music composed to facilitate high arousal experiences can incorporate more complex ternary rhythmic pattern, fast tempos, bright timbres, staccato articulations, complex musical textures, as well as unexpected or novel musical events (e.g. a sudden rhythmic change). Music composed to facilitate low arousal experiences can incorporate soft dynamics, a low-than-normal pitch range, slow tempos, ritardandos, simple musical textures, and legato articulations. These theory-based guidelines informed the composition of the music included in the MCRF intervention, which is currently being evaluated for fidelity.

ISBN 1-876346-65-5 © ICMPC14

Representing change in patterns of interaction of children with communication difficulties in music therapy: what do the therapists think?

Neta Spiro[1], Camilla Farrant[1] and Tommi Himberg[2]
[1]Research Team, Nordoff Robbins, London, UK
[2]Department of Neuroscience and Biomedical Engineering, Aalto University, Espoo, Finland

Music therapy has been found to help improve communicative behaviours and joint attention in children with communication difficulties, but it is unclear what processes in the music therapy sessions contribute to those changes. Previous research has focused on the use of outcome measures or on the analysis of segments of sessions. One way to explore what contributes to change is to analyse what happens in the sessions. In developing methods for such analysis, it is informative to understand how the therapists in the sessions view the results of the analysis. In this two-stage project, we created representations of aspects of interactive behaviours in improvisational music therapy sessions. We then asked the therapists who had been in the videos for their views about the representations. To create the representations, an annotation protocol was developed which focused on simple, unambiguous individual and shared behaviours of clients and therapists including, facing, rhythmic activity and musical structures. We analysed sessions with children who had been referred for communication difficulties. The representations were of five pairs of whole early and late sessions and showed either moment-by-moment views of behaviours of each player or proportions of the annotated behaviours. Three of the therapists who had been in the videos responded to a questionnaire about the representations. They reported that all the characteristics presented were relevant to tracing change. They found that the visualisations would be useful for their own assessments of client progress as well as for communication about their work with others. The role of these representations was discussed in the context of a balance between consideration of whole sessions and the analysis of salient moments within sessions. The therapists' responses indicate that though the representations illustrate simple behaviours, they may contribute to the analysis of the multi-faceted work of the music therapists.

A two-year longitudinal investigation comparing pitch discrimination and pitch matching skills of young children

Debbie Lynn Wolf[1]
[1]School of Music, Cairn University, Langhorne, Pennsylvania, USA

Many researchers have explored relationships between the ability to discriminate pitch and the ability to match pitch, but results conflict particularly for young children. Although tonal patterns may be more accessible for children to match than songs, most researchers have focused on cross-sectional investigations of song reproductions. A longitudinal study of discrimination and pitch matching of tonal patterns may provide a more equitable comparison and greater insight into children's initial musical experiences.

The purpose of this study was to examine longitudinal differences and relationships between children's pitch discrimination and pitch matching skills.

The researcher administered a standardized tonal aptitude test as a discrimination measure and an investigator-designed imitative response test as a pitch matching measure to 96 students at the end of kindergarten and second grade.

Differences in pitch discrimination and pitch matching between kindergarten and second grade scores were examined by paired t-tests. Results indicate improvement of 14.5% in discrimination and 13% in pitch matching between kindergarten and second grade. Pearson correlations between the score distributions of the discrimination measure and the pitch matching measure for both grades were analyzed. The relationship between discrimination and pitch matching remains moderate in kindergarten (.40) through second grade (.45). The strong relationship between kindergarten and second grade pitch matching scores (.74) contrasts with the moderate relationship between kindergarten and second grade discrimination scores (.53).

Significant improvement in pitch discrimination and pitch matching skills can be expected to occur at similar rates between kindergarten and second grade. Pitch matching skills seem to stabilize prior to discrimination skills. Findings suggest early elementary music instruction may have greater impact on discrimination than pitch matching skills.

How Masculine Is a Flute?
A Replication Study on Gender Stereotypes and Preferences for Musical Instruments among Young Children

Claudia Bullerjahn,*1 Katharina Heller,*2 Jan H. Hoffmann*3

*Institute for Musicology and Music Education, Justus-Liebig-University Giessen, Germany

1Claudia.Bullerjahn@musik.uni-giessen.de, 2Katharina.Heller@musik.uni-giessen.de, 3Jan.H.Hoffmann@musik.uni-giessen.de

ABSTRACT

Our study aimed at replicating the previous findings of Harrison and O'Neill (2000) with a younger sample (N = 90, aged 4 to 6) and expanding the repeated measures design to detect the influence of different treatments on preference as well as gender stereotypes. During a priming phase, all children were accustomed with look and sound of six musical instruments. In the following intervention phase, groups 1 and 2 received concerts played by either gender-consistent or gender-inconsistent role models and group 3 received an episode of a TV-series for kids. Regarding both boys and girls, results reveal no significant changes in terms of instrumental preferences except for the guitar (♂: p = .026, ♀: p = .001): all subjects tend to prefer this instrument more at the second time of measurement. Nevertheless, differences among the three groups are not significant in terms of preferences. Regarding gender stereotypes, boys (p = .001) as well as girls (p = .061) show significant differences or tendencies for flute. With regard to drums, girls show a significant interaction (p = .031) in reference to time and group. Nevertheless, post hoc tests do not indicate any significance among the three different groups. Regardless of sex, children assign all instruments rather according to their own gender. Our findings clearly differ from the previous, perhaps caused by confounding variables and methodological problems such as children's conversations between priming and intervention phase. However, we think that our results are mainly due to the fact that today's boys and girls have developed distinguished and surprisingly stable preferences for musical instruments even at this young age. Contacts beyond the interventions may explain the observed changes in preferences for the guitar. Nevertheless, gender identities and gender-typed activities correspond with social contexts and therefore each instrument can be appropriate, what offers best chances for early instrumental learning.

I. INTRODUCTION

> *Concern has been increasingly expressed about boys not participating as fully as they might in musical activities, and it is recognized that they do not have the same wide range of interests and preferences as girls have for learning musical instruments. The question that arises is whether boys are actually being excluded due to girls' greater participation, or that their more limited range of interest for playing certain instruments is a result of the influence of gender associations.* (Bruce & Kemp, 1993, p. 213).

Is gender decisive for the musical career? Women's movements and other social and political actions have called for equal opportunities for males and females in the workplace or at schools and should have diminished the influence of role models. However, associations between gender and specific musical instruments still remain today: Wrape, Dittloff, and Callahan (2016) examined gender stereotypes of band instruments among public middle school students and found out that tuba, percussion and trombone were categorized as boy instruments, while clarinet, flute and oboe were particularly associated with girls. Zervoudakes and Tanur (1994) concluded after the examination of concert programs over three decades that the increase in females playing typically male instruments was not so much a function of reduction in gender stereotypes as an overall increase in female participation. Talking about gender stereotypes in this study means that instruments are afflicted with prejudices such as being 'typically male' or 'typically female'. The sex stereotyping of musical instruments is a problem, because it tends to limit the range of musical experiences available to male and female musicians in several ways, including participation in instrumental ensembles and playing in specific musical styles.

There are numerous studies that explored the relationship between musical instruments and their association with a specific gender (e.g. Abeles & Porter, 1978; Griswold & Chroback, 1981; Delzell & Leppla, 1992; Fortney, Boyle, & DeCarbo, 1993; Tarnowski, 1993; Zervoudakes & Tanur, 1994; O'Neill & Boulton, 1996; Elliot & Yoder-White, 1997; Sinsel, Dixon, & Blades-Zeller, 1997; Conway, 2000; Harrison & O'Neill, 2000; Pickering & Repacholi, 2001; Cramer, Million, & Perreault, 2002; Harrison & O'Neill, 2003; Graham, 2005; Abeles, 2009; MacLeod, 2009; Payne, 2009; Killian & Satrom, 2011; Marshall & Shibazaki, 2011; Vickers, 2015; Wrape, Dittloff, & Callahan, 2016; cf. for an overview Eros, 2008 and Wych, 2012). All found "gender biases both in what instruments children choose to play and their attitudes about what others should play" (Wrape et al., 2016, p. 40). Participants were typically asked to rate their preference for a selection of musical instruments or to place an instrument along a feminine – masculine continuum according to its perceived association with gender. In most studies, flute and violin have been the feminine anchors, while drums and trumpet anchored the masculine end (Abeles & Porter, 1978; Griswold & Chroback, 1981; Delzell & Leppla, 1992; Abeles, 2009).

Several researchers are of the opinion that observational learning may be responsible for establishing and fostering gender stereotypes (Cramer, Million, & Perreault, 2002; Delzell & Leppla, 1992; Fortney, Boyle, & DeCarbo, 1993). Nonetheless, gender atypical instrument modeling as an intervention to counter gender-stereotyped instrument selections often has only weak effects (Bruce & Kemp, 1993; Harrison & O'Neill, 2000; Pickering & Repacholo, 2001; Killian & Satrom, 2011; Vickers, 2015).

The study of Anna Harrison and Susan O'Neill (2000) served as basis for the present survey. They investigated preferences for musical instruments as well as gender stereotyped associations of those instruments with children aged seven to eight ($N = 357$). Their study aimed to examine, whether gender-typed preferences will either be established, confirmed or changed after exposure to stereotypic or counter-stereotypic role models. During a priming phase, six instruments (flute, violin, piano, trumpet, guitar and drums) were introduced at schools in Great Britain. Short interviews with pictures of the instruments were held and preferences as well as gender stereotyped associations to those instruments were captured. An intervention concert followed this phase. Subjects were divided into three groups: Group 1 and group 2 received concerts played by either gender-consistent or gender-inconsistent role models and group 3 served as control group. It is important to take into account that Harrison and O'Neill mention no alternative treatment for the control group.

Results confirmed earlier findings regarding role models and preferences. Girls showed a significant change concerning preferences for flute, piano and drums. Girls who received a gender-inconsistent concert rejected the piano more after the concert. A similar but not significant trend could be found for flute and violin. Boys in the two main groups (gender-consistent and -inconsistent) preferred the drums more after the concert. In addition, only the results for violin showed a tendency for rejection after seeing a female play the instrument (cf. Harrison & O'Neill, 2000).

We choose a replication of this study, especially because of the innovative research design with live concerts. Throughout the last decades, research regarding the preference for a particular musical instrument as well as gender stereotypes with children and adolescents were more or less limited to the United States and Great Britain. In Germany, there are only very few studies exploring the relationship between musical instruments and gender (Scheuer, 1988; Vogl, 1993; Müller & Burr 2004). Since stereotypes and preferences are established much earlier than at the age of seven, the present study investigates effects among younger children.

II. RESEARCH ISSUES

Besides capturing preferences for a particular musical instrument as well as gender stereotypes, we aspire to examine the influence that a performance involving the instruments in question has on exactly those variables. Our aim is to ascertain whether the sex of a musician has an impact on existing preferences for musical instruments and gender stereotypes. Moreover, what happens when existing role models are met or broken should be examined.

III. MATERIALS AND METHODS

A. Sample

The sample comprises a group of young German children ages four to six (♂ = 48; ♀ = 42; $N = 90$), on an average they are five years old ($M = 5.04$ $SD = .669$). All subjects attend one of five different day care centers. For the examination, all children were divided into two experimental groups – the gender-consistent group (group 1: $n = 26$) and the gender-inconsistent group (group 2: $n = 37$) –, and one control group (group 3: $n = 27$). Families own at least one musical instrument in 70 % of the present cases. 44.4 % of the participants were already playing this instrument at the time of survey. In 26 cases, the child was the only one to play an instrument in their family. 69.7 % of the young children had already experienced live music in a concert or a similar situation. For the remaining 30.3 % the concert experienced during the examination seems to be the first live music performance.

B. Data collection instrument

We use a specifically developed questionnaire as the data collection instrument for the present examination. It was limited to few items to ensure that our young subjects are able to complete it in an appropriate amount of time and with the assistance of the interviewer. Content and amount were discussed in a colloquium with teachers and educators. Upon consultation, the processing time was set for 15-20 minutes.

The first part of the questionnaire aimed to collect socio-demographic data to be used as general exploratory information. The second part of the questionnaire compiled preferences for particular musical instruments as well as gender stereotypes. The included instruments were flute, violin, piano, trumpet, guitar and drums. Because the children were too young to read, continuous scales and drawings were used to rate the instruments for preferences as well as for stereotypes. First, the children should place small pictures of the six musical instruments on a rectangular field in rank order according to their preference for the instruments (the closer to the upper rim, the better liked; placed side by side means same preference). Second, they should make one's cross on lines with arrows on ends pointing towards a drawing of a boy or a girl for every instrument (the nearer the cross to the respective drawing, the more likely it should be played by the respective gender, see figure 1). To code the data, we measured out the place of the pictures or the crosses and applied a metrical splitting of the continuums.

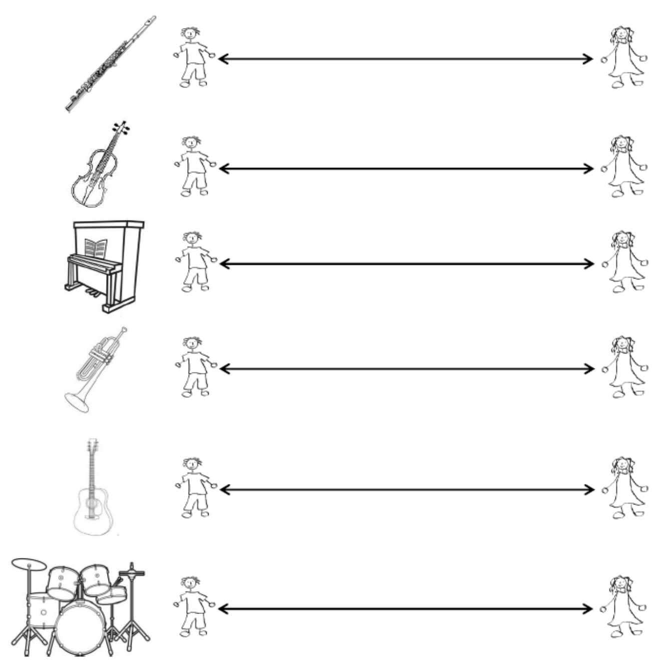


Figure 1. Third page of the questionnaire with continuous rating scales for estimating, whether a boy or a girl should rather play a musical instrument.

C. Procedure

The time sequence divided into two survey phases and each survey phase into two sub-phases. The first stage of the survey consisted of a priming phase followed by the first questioning (t1). During the priming phase, the researchers wanted to ensure that all of the children had at least a first impression of the presented instruments. To avoid gender stereotypes and traditional role models, the subjects saw pictures of each instrument on a white background. Additionally, each playing technique was explained and sound recordings (interpretations of the melody *Twinkle, Twinkle Little Star* and a C major scale; for the drums a short drum pattern) were played. The priming phase took approximately 20 minutes and was carried out in smaller groups of five children. The second stage of the survey took place one week after the first. It consisted of the respective intervention concert, immediately followed by a second round of questioning (t2).

The two experimental groups underwent slightly different treatments. Male and female musicians presented live one of six different instruments each to all of these children. In the group 1, the sex of the respective musician complies with the stereotypic role models which were confirmed by earlier studies (Abeles & Porter, 1978; Bussey & Perry, 1982; Delzell & Leppla, 1992; Bruce & Kemp, 1993). We defined this group as the gender-consistent group. The opposite cast (e.g. a male playing the flute and a female playing the drums) was presented to the group 2 – the gender-inconsistent group. Group 3 served as the control group. The children of this group received an episode of a TV-series for kids.

The intervention concerts took place seven days after the first ascertainment. We presented the musical instruments in random order. Each instrument played a short solo-piece. Timing and musical style were similar in both groups.

D. Statistical Methods

All data was analyzed using the *Statistical Packages for the Social Sciences* (*SPSS*, Version 22). Two-way repeated measures variance analyses (ANOVA) between three groups computed for both gender were used for answering the research issues as well as t-tests for independent samples. The level of significance was set at $\alpha = .05$.

IV. RESULTS

Right after the priming phase (t1), boys indicated highest preferences for the drums ($M = 1.36$) and lowest for the flute ($M = 4.38$), whereas girls preferred most the piano ($M = 2.11$) and least the trumpet ($M = 4.94$) (cf. table 1). The difference between the preference ratings of boys and girls is significant for the trumpet ($F = 4.411$, $p = .015$) and nearly significant for the drums ($F = 2.874$; $p = .062$). These findings are in line with gender-typical preferences measured in previous studies (e.g. Abeles & Porter, 1978; Griswold & Chroback, 1981; Delzell & Leppla, 1992; O'Neill & Boulton, 1996; Harrison & O'Neill, 2000; Abeles, 2009; Killian & Satrom, 2011; Vickers, 2015). It is remarkable that the children rejected no musical instrument at the first time of measurement.

Table 1. Boys and girls' preferences for musical instruments right after the priming phase (t1).

		flute	violin	piano	trum-pet	guitar	drums
boys	*M*	4.375	3.746	3.158	4.223*	2.948	1.369(*)
	SD	3.301	3.739	3.150	3.311	3.084	2.154
girls	*M*	3.067	2.926	2.110	4.940*	3.812	3.352(*)
	SD	3.338	3.076	2.322	3.370	3.253	3.442

M: mean-value (0 = absolutely preferred, 9 = absolutely rejected); *SD*: standard deviation; one-way ANOVA: (*) p ≤ .10; * p ≤ .05; ** p ≤ .01; *** p ≤ .001

In contrast to our expectations, children assigned all instruments rather according to their own gender (cf. table 2): Whereas girls stated that all musical instruments except the trumpet and the drums should rather be played by a girl with the piano as most feminine ($M = 8.42$) and the trumpet as most masculine ($M = 4.96$), boys categorized all musical instruments more or less as boy instruments with the drums as most masculine ($M = 2.06$) and the piano as least masculine ($M = 5.55$). There is only one significant difference between boys and girls for the trumpet ($F = 3.519$, $p = .034$).

Table 2. Gender stereotypes for musical instruments right after the priming phase (t1).

		flute	violin	piano	trum-pet	guitar	drums
boys	*M*	4.375	4.635	5.552	3.479*	2.583	2.063
	SD	4.262	4.607	4.367	3.771	3.428	3.212
girls	*M*	8.369	7.738	8.417	4.964*	6.583	5.583
	SD	3.957	4.427	3.641	4.426	4.614	4.628

M: mean-value (0 = absolutely masculine, 12 = absolutely feminine); *SD*: standard deviation; one-way ANOVA: (*) p ≤ .10; * p ≤ .05; ** p ≤ .01; *** p ≤ .001

Concerning both boys and girls, results reveal no significant differences between the two times of measurement in terms of preferences for musical instruments except for the guitar (see table 3).

Table 3. Preference for the guitar: two-way analysis of variance (ANOVA) with repeated measures between the three groups computed for both gender.

		F	*df*	*p*	*d*
boys	time	5.292	1 / 45	.026	0.456
	time*group	.198	2 / 45	.821	
girls	time	11.788	1 / 39	.001	0.598
	time*group	1.372	2 / 39	.266	

F, F-value; *df*, degrees of freedom; *p*, empirical significance; *d*, Cohen's d (effect size) with: 0.00–0.10 (no effect), 0.20–0.40 (small), 0.50–0.70 (medium), > 0.80 (large).

All subjects tend to prefer the guitar more at the second time of measurement (see figure 2 and 3).

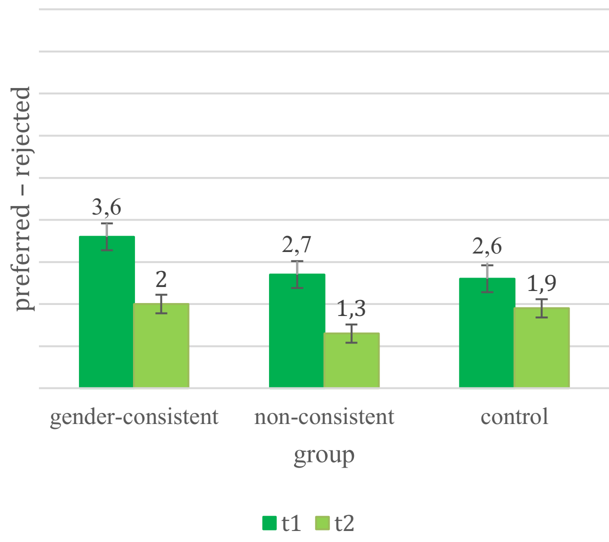


Figure 2. Boys' preference for the guitar: comparison of the two times of measurement between the three groups.

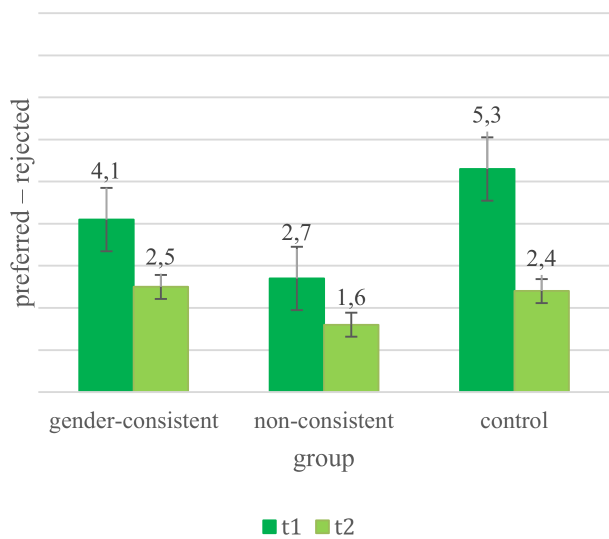


Figure 3. Girls' preference for the guitar: comparison of the two times of measurement between the three groups.

Regarding preferences for all instruments, the differences between the three groups are not significant.

Considering gender stereotypes, two instruments show significant results or at least tendencies between the two times of measurement: flute for boys and girls, violin for boys only. With regard to drums, girls show a significant interaction in reference to time and group (see table 4). Nevertheless, post hoc tests do not reveal any significance between the three different groups.

Table 4. Gender stereotypes for different instruments: two-way analysis of variance (ANOVA) with repeated measures between the three groups computed for both gender.

			F	*df*	*p*	*d*
boys	flute	time	12.152	1 / 45	.001	0.617
		time*group	2.428	2 / 45	.100	
	violin	time	3.172	1 / 45	.082	
		time*group	.909	2 / 45	.410	
girls	flute	time	3.721	1 / 39	.061	
		time*group	.780	2 / 39	.465	
	drums	time	.673	1 / 39	,417	
		time*group	3.812	2 / 39	.031	

F, F-value; *df*, degrees of freedom; *p*, empirical significance; *d*, Cohen's d (effect size) with: 0.00–0.10 (no effect), 0.20–0.40 (small), 0.50–0.70 (medium), > 0.80 (large).

Interestingly, both boys and girls rate the flute in the same direction but with distinct differences. Boys tend to rate the flute as more 'masculine' at first, and more 'feminine' after the second time of measurement; whereas, girls also tend to rate the flute as more 'feminine' over time as well, but already originally see it as more 'feminine' than boys do in the beginning at first without intervention (see figure 4). A t-test for independent samples reveals that this difference between boys and girls is significant (t (90) = -4.59, p ≤ .001). This result is independent from any intervention that took place.

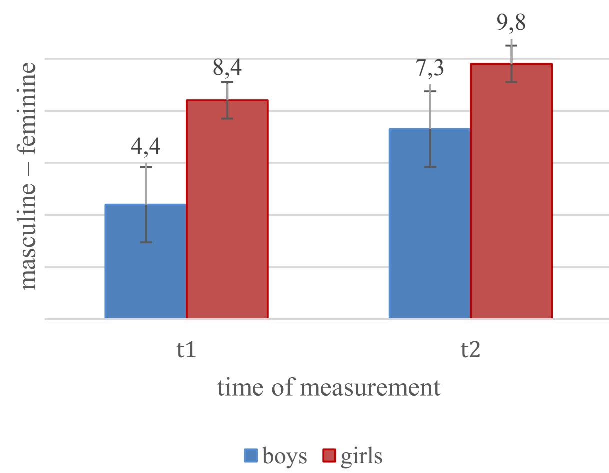


Figure 4. Gender stereotypes for the flute between two times of measurement.

V. DISCUSSION

Are preferences for musical instruments trending towards change or are they long-standing in terms of gender-stereotypes? Our results show that today's boys and girls have developed distinguished and surprisingly stable preferences for musical instruments even at this young age – may be due to parental or media influences. They correspondent broadly similar in direction to findings by Scheuer (1988), Fortney et al. (1993), O'Neill & Boulton (1996), Conway (2000), Müller & Burr (2004), Pickering & Repacholi (2001); Marshall & Shibazaki, 2011 and Killian & Satrom (2011). Therefore, our results confirm the highly prevalent stereotypes in musical instrument selection and attitudes on musician gender and instrumentation in general, but we would like to emphasize above this that the participating children did not reject a musical instrument particularly and the standard deviations were rather high. The results in study 3 by Abeles & Porter (1978) indicated something similar: the sex-stereotyping behaviour in musical instrument preference was not very strong in kindergarten children but more pronounced in children beyond school grade 3. However, our findings might be also due to the new developed measuring method with continuums instead of forced-choice categories. In addition, we did not ask the children to select the instrument that they would like to play, but to rank order all instruments in terms of preference. Because they should perform choices without real consequences (e.g. the opportunity to play the instrument liked best), they might not be afraid of negative outcomes such as playing a gender-inappropriate instrument in the presence of peers.

But yet, our results clearly differ from the previous in terms of changes between the two times of measurement what might be also caused by confounding variables and methodological problems such as children's conversations between priming and intervention phase. Contacts with the guitar beyond the interventions can explain the only observed changes in preferences for this instrument, even though we do not know of any special contacts.

Children establish a gender identity latest in the third year of life and preschoolers typically show the strongest relations among gender schemata. However, at the end of preschool and with the beginning of elementary school an increase in stereotype flexibility is frequently observed which might signal that the interrelations among gender schemas are weakened (Signorella, 1999; Zosuls, Miller, Ruble, Martin, & Fabes, 2011). In addition, gender-typed activities vary in strength dependency upon their social contexts (Goble, Martin, Hanish, & Faves, 2012) and the children in our study were asked in smaller groups. Moreover, the presence of instrumental stereotypes is equivocal in preschool children (Abeles & Porter, 1978; Marshall & Shibazaki, 2012). All this can explain that in our study each musical instrument is perceived as appropriate for one's own sex. But yet in a study on self-perceived gender typicality and gender stereotypes in general, Meagan Patterson (2012) could show "that even young elementary-school-aged children used their knowledge of cultural gender roles to make subjective judgments regarding the self, and, conversely, that views of the self may influence personal endorsement of cultural gender stereotypes" (p. 422). In accord with Wrape et al. (2016) it could be confirmed that our young and musically inexperienced children are more open to counter-stereotypical views than older children. This offers best chances for early instrumental learning.

ACKNOWLEDGMENT

The authors would like to thank Lydia Scott for additional proofreading.

REFERENCES

Abeles, H. F., & Porter, S. Y. (1978). The Sex Stereotyping of Musical Instruments. *Journal of Research in Music Education, 26* (2), 65-75.

Abeles, H. (2009). Are Musical Instrument Gender Associations Changing? *Journal of Research in Music Education, 57* (2), 127-139.

Bruce, R., & Kemp, A. (1993). Sex-stereotyping in Children's Preferences for Musical Instruments. *British Journal of Music Education, 10* (3), 213-217.

Bussey, K., & Perry, D. G. (1982). Same-Sex Imitation: The Avoidance of Cross-Sex Models or the Acceptance of Same-Sex Models? *Sex Roles, 8* (7), 773-784.

Conway, C. (2000). Gender and musical instrument choice: A phenomenological investigation. *Bulletin of the Council for Research in Music Education, 146*, 1-15.

Cramer, K. M., Million, E., & Perreault, L. A. (2002). Perceptions of musicians: Gender stereotypes and social role theory. *Psychology of Music, 30*, 164-174.

Delzell, J. K., & Leppla, D. A. (1992). Gender Association of Musical Instruments and Preferences of Fourth-Grade Students for Selected Instruments. *Journal of Research in Music Education, 40*, 93-103.

Elliot, C. A., & Yoder-White, M. (1997). Masculine/feminine associations for instrumental timbres among children seven, eight, and nine years of age. *Contributions to Music Education, 24* (2), 30-39.

Eros, J. (2008). Instrument Selection and Gender Stereotypes. A Review of Recent Literature. *Update: Applications of Research in Music Education, 27* (1), 57-64.

Fortney, P. M., Boyle, J. D., & DeCarbo, N. J. (1993). A study of middle school band students' instrument choices. *Journal of Research in Music Education, 41*, 28–39.

Goble, P., Martin, C. L., Hanish, L. D., & Faves, R. A. (2012). Children's Gender-Typed Activity Choices across Preschool Social Contexts. *Sex Roles, 67*, 435-451.

Graham, B. J. (2005). *Relationships among instrument choice, instrument transfer, subject sex, and gender-stereotypes in instrumental music*. Dissertation. Indiana University: ProQuest (UMI No. 3298144).

Griswold, P. A., & Chroback, D. A. (1981). Sex-role associations of music instruments and occupations by gender and major. *Journal of Research in Music Education, 29*, 57–62.

Hallam, S., Rogers, L., Creech, A. (2008). Gender differences in musical instrument choice. *International Journal of Music Education, 26* (1), 7-19.

Harrison, A. C., & O'Neill, S. A. (2000). Children's Gender-Typed Preferences for Musical Instruments. An Intervention Study. *Psychology of Music, 28*, 81-97.

Harrison, A. C., & O'Neill, S. A. (2003). Preferences and Children's Use of Gender-Stereotyped Knowledge about Musical Instruments: Making Judgments about Other Children's Preferences. *Sex Roles, 49* (7/8), 389-400.

Killian, J. N., & Satrom, S. L. (2011). The effect of demonstrator gender on wind instrument preferences of kindergarten, third-grade, and fifth-grade students. *Update: Applications of Research in Music Education, 29* (2), 13-19.

MacLeod, R. B. (2009). A comparison of aural and visual instrument preferences of third and fifth-grade students. *Bulletin of the Council for Research in Music Education, 179*, 33-43.

Marshall, N. A., & Shibazaki, K. (2011). Instrument, Gender and Musical Style Associations in Young Children. *Psychology of Music, 40* (4), 494-507.

Müller, R., & Burr, M. (2004). Präsentative Forschungsmethoden zur Untersuchung von Musikinstrumentenpräferenzen in Schulen. In B. Hofmann (Ed.), *Was heißt methodisches Arbeiten in der Musikpädagogik?* (pp. 149-168). Essen: Die Blaue Eule.

O'Neill, S. A., & Boulton, M. J. (1996). Boys' and Girls' Preferences for Musical Instruments: A Function of Gender? *Psychology of Music, 24*, 171-183.

Patterson, M. M. (2012). Self-Perceived Gender Typicality, Gender-Typed Attributes, and Gender Stereotype Endorsement in Elementary-School-Aged Children. *Sex Roles, 67*, 422-434.

Payne, P. D. (2009). *An Investigation of Relationships between Timbre Preference, Personality Traits, Gender, and Music Instrument Selection of Public School Band Student*. Dissertation. The University of Oklahoma: ProQuest (UMI No. 3366051).

Pickering, S., & Repacholi, B. (2001). Modifying Children's Gender-Typed Musical Instrument Preferences: The Effects of Gender and Age. *Sex Roles, 45* (9/10), 623-643.

Scheuer, W. (1988). *Zwischen Tradition und Trend. Die Einstellung Jugendlicher zum Instrumentalspiel. Eine empirische Untersuchung*. Mainz: Schott.

Signorella, M. L. (1999). Multidimensionality of gender schemas: Implications for the development of gender-related characteristics. In W. B. Swann Jr., J. H. Langlois & L. A. Gilbert (Eds.), *Sexism and stereotypes in modern society: The gender science of Janet Taylor Spence* (pp. 107-126). Washington: American Psychological Association.

Sinsel, T. J., Dixon, W. E., & Blades-Zeller, E. (1997). Psychological Sex Type and Preferences for Musical Instruments in Fourth and Fifth Graders. *Journal of Research in Music Education, 45* (3), 390-401.

Tarnowski, S. M. (1993). Gender-bias and musical instrument preference. *Update: Applications of Research in Music Education, 12* (1), 14-21.

Vickers, M. E. (2015). *The effect of model gender on instrument choice preference of beginning band students*. Dissertation. University of Hartford – The Harrt School: ProQuest (UMI No. 3700787).

Vogl, M. (1993). *Instrumentenpräferenz und Persönlichkeitsentwicklung. Eine musik- und entwicklungspsychologische Forschungsarbeit zum Phänomen der Instrumentenpräferenz bei Musikern und Musikerinnen*. Frankfurt/M.: Peter Lang.

Wrape, E. R., Dittloff, A. L., & Callahan, J. L. (2016). Gender and Musical Instrument Stereotypes in Middle School Children: Have Trends Changed? *Update: Applications of Research in Music Education, 34* (3), 40-47.

Wych, G. M. F. (2012). Gender and Instrument Associations, Stereotypes, and Stratification: A Literature Review. *Update: Applications of Research in Music Education, 30* (2), 22-31.

Zervoudakes, J., & Tanur, J. M. (1994). Gender and musical instruments: Winds of change? *Journal of Research in Music Education, 42*, 58-67

Zosuls, K. M., Miller, C. F., Ruble, D. N., Martin, C. L., & Fabes, R. A. (2011). Gender Development Research in *Sex Roles*: Historical Trends and Future Directions. *Sex Roles, 64*, 826-842.

The importance of musical context on the social effects of interpersonal synchrony in infants

Laura Cirelli[1], Stephanie Wan[1], Christina Spinelli[1], Laurel Trainor[1,2,3]
[1]Department of Psychology, Neuroscience & Behaviour, McMaster University, Canada
[2]McMaster Institute for Music and the Mind, McMaster University, Canada
[3]Rotman Research Institute, Baycrest Hospital, Toronto, Ontario, Canada

Interpersonal synchrony encourages prosocial behavior among adults, children, and even infants. In adults, this effect has been demonstrated in both musical and non-musical contexts. Music is therefore considered to be a natural, but not necessary, context within which interpersonal synchrony is achieved. However, how the presence of music shapes the social effects of interpersonal synchrony has yet to be directly addressed. Previous work in our lab has shown that 14-month-old infants are more helpful towards an experimenter after being bounced synchronously versus asynchronously with that person. In these experiments, music was always present in the background. This experiment investigated the effects of interpersonal synchrony and asynchrony on infant helping in a non-musical context. As in our previous experiments, infants were held by an assistant and bounced at a steady rate. At the same time, the experimenter facing the infant bounced either in- or out-of-synchrony with how the infant was bounced. In contrast to our previous studies in which infants listened to music while being bounced, here infants listened to nature sounds (i.e. rushing water, rustling leaves). Afterward, the infant was placed in a situation where the experimenter attempted to complete simple tasks, but dropped the objects she needed to complete the tasks. The number of dropped objects handed back by each infant was recorded, and average helping rates were compared across conditions. Replicating our previous findings, infants helped significantly more in the synchronous versus asynchronous conditions. However, infants displayed more distress in this experiment (40% did not complete the experiment compared to our typical rates of 15-25%) and infants took longer to help than in our previous experiments. These results suggest that while music may not be necessary for the social effects of synchrony to emerge, the presence of music does contribute positively to the experience.

Melody familiarity facilitates music processing in 4-5-year-olds

Sarah C. Creel[*1]

[*]Department of Cognitive Science, UC San Diego, USA
[1]screel@ucsd.edu

ABSTRACT

How does music recognition develop? One possibility is that children begin by learning a small "lexicon" of highly-frequent melodies, which eventually develop into a full musical system. This account is inspired in part by findings from language acquisition: in toddlers, familiar words (dog; Swingley & Aslin, 2002) show more detailed sound representations than unfamiliar words (dih; Stager & Werker, 1997). The role of familiarity in music acquisition has received some attention (Corrigall & Trainor, 2010), but more work is needed. The current study assessed whether familiar melodies show facilitated processing in young children (ages 3-5 years, N=32 per experiment). Children in each experiment heard either familiar melodies or scrambled versions of the same melodies. Experiment 1 asked children to associate each melody with a picture. Children were significantly better at associating familiar melodies with pictures than scrambled melodies. However, the familiar-melody advantage only occurred for familiar-picture associates (star, lamb), not unfamiliar pictures (two cartoon animals). This mirrors findings from language learning. Experiment 2 asked children to say whether a pair of melodies were the same or different. They were significantly better at distinguishing the familiar melodies than the unfamiliar melodies, suggesting that melody familiarity facilitates maintenance in auditory short-term memory. At a theoretical level, children's facilitated processing of familiar melodies suggests the beginnings of a process of musical enculturation via lengthy encoding of musical exemplars, mirroring the emergence of phonological systems in spoken languages by acquisition of vocabulary.

I. INTRODUCTION

How do children represent musical events? Much literature in music perception focuses on child (and adult) apprehension of abstract musical constructs such as scale structure and meter. This focus might be interpreted as an assumption that scale degrees form the "alphabet," or representational primitives, out of which melodic representations are constructed. A different view is that musical representations are stored as detailed events (see Goldinger, 1998, for such an account of language representations), with abstractions like scale structure, probabilistic melodic patterns, and meter then emerging secondarily out of a large collection of events. This second view parallels accounts of language acquisition (Feldman, Myers, White, Griffiths, & Morgan, 2013): rather than acquiring speech sounds first and only then acquiring words, words are learned first, and speech sounds and probabilistic speech-sound combinations emerge out of (or by analogy to) known words. The current study explores whether this specific learning hypothesis applies to music. If so, then children should be better at processing familiar musical events than unfamiliar ones.

In the domain of word learning, word knowledge appears to be critical to speech sound processing. Young children (14-month-olds) can *distinguish* similar speech sounds such as /b/ and /d/, but they perform poorly at learning labels distinguished by these words ("bih" and "dih") at the same age, only succeeding at 17-20 months (Stager & Werker, 1997; Werker, Fennell, Corcoran, & Stager, 2002). However, *familiar* words that are distinguished by similar speech sounds (ball and doll) are readily recognized by children aged 14-15 months (Fennell & Werker, 2003; Swingley & Aslin, 2002). This suggests that memory for speech sounds in unfamiliar contexts may be quite poor, but memory for the same speech sounds in familiar words is good.

A parallel set of findings from music perception is that preschool-aged children can auditorily discriminate musical sounds that they cannot associate with visual objects (cartoons; Creel, 2014, 2016). However, their discrimination of sounds is also less accurate for the hard-to-associate sounds (Creel, 2014, 2016). This is especially surprising in that the hard-to-associate sounds are those that differ in pitch contour, which is generally regarded as developmentally early (Trehub, Bull, & Thorpe, 1984) and as being central to melodic identity (Dowling, 1978).

Some previous research supports the idea that children learn specific musical events rather than abstract musical representations. Two studies by Trainor and colleagues suggest that 4-year-olds are less sensitive to musical structure violations for unfamiliar melodies (Trainor & Trehub, 1994) than for familiar melodies, specifically, Twinkle Twinkle Little Star (Corrigall & Trainor, 2010). However, no such effects have been tested within a single experiment.

One possible explanatory basis for these findings in speech and in music in terms of the specific learning hypothesis is that children are accruing *specific memories* of musical (and linguistic) patterns, such as melodies and words, rather than learning more abstract structures (major scale, phonology). Thus, relatively weaker performance with less-familiar materials is driven by weaker underlying auditory memory representations (of the unfamiliar nonsense word "bih" or an unfamiliar melody) relative to more-familiar representations (of the familiar word "ball" or a familiar melody like Twinkle Twinkle Little Star). On this **specific learning hypothesis**, more-familiar items should be both *easier to discriminate* and *easier to associate*.

To assess the specific-learning hypothesis, the current study tested preschool-aged children's abilities to discriminate and to form audiovisual associations with either familiar melodies or scrambled (thus unfamiliar) versions of those melodies.

II. Pretesting

Initial pilot work aimed to determine what melodies children were best at identifying, out of a set of likely candidates. The first pretest tested 19 preschool-aged children by presenting either two or four measures of six childhood melodies (Mary Had a Little Lamb, Twinkle Twinkle Little Star, Happy Birthday, Deck the Halls, Yankee Doodle, and London Bridge), played in a piano timbre, and asking children to name them. Melodies will be referred to hereafter with only the first word of the title for brevity. Naming accuracy was fairly low (21% overall) and did not differ between 2-measure and 4-measure melodies. The highest naming rates were for Happy (36%), Mary (30%), and Twinkle (39%).

Table 1. Pretest stimuli.

Melody	Picture
Mary Had a Little Lamb	lamb
Twinkle Twinkle Little Star	star
Happy Birthday	cake with candles
Deck the Halls	Christmas tree
Yankee Doodle	USA Revolutionary soldier
London Bridge	(not used in pretest 2)
Star-Spangled Banner	flag

A second pretest with 23 children used a 2-alternative forced-choice picture-matching task with the same melodies, except than London Bridge (0% naming recognition) was replaced with the Star-Spangled Banner. Pictures were Here, overall accuracy was 59%. Length (2 vs. 4 measures) did not impact recognition. The pairs with highest accuracy were Mary-Twinkle and Birthday-Twinkle (both 69.6% for the 2-measure length). The two chosen melodies were Mary and Twinkle because they have similar rhythmic patterns. After selecting these two melodies, scrambled versions of each were created, attempting to use simple contours for both (Figure 1) so that familiar and unfamiliar melodies were matched for overall naturalness.

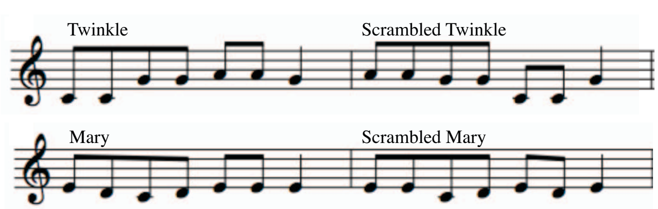


Figure 1. Experiments 1 and 2, melody stimuli.

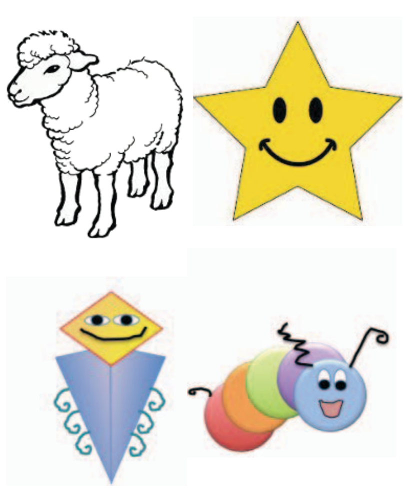


Figure 2. Experiment 1, visual stimuli. Top row: familiar pictures. Bottom row: novel pictures.

III. Experiment 1: Association learning

A. Method

1) *Participants*. Sixty-four preschool-aged children (ages 3-5) took part.

2) *Stimuli*. The four melodies were Mary, Twinkle, Mary-scrambled, and Twinkle-scrambled. Two different picture sets were used (Figure 2): either a lamb and a star (familiar-picture condition), or two cartoon characters (unfamiliar-picture condition).

3) *Procedure*. Children received reinforced learning trials in blocks of 8 trials each (4 with one melody-picture combination, 4 with the other). On each trial, two pictures appeared on the left and the right side of the screen, and then a melody played. The child was asked to select the picture that "goes with" the melody. Once they scored at least 7/8 in a block of learning trials, they continued to unreinforced test trials. Sixteen children each learned to associate familiar melodies with familiar pictures; unfamiliar melodies with familiar pictures; familiar melodies with unfamiliar pictures; unfamiliar melodies with unfamiliar pictures.

B. Results and Discussion

Results from testing trials suggested that performance degraded when feedback disappeared, so are not analyzed further. Results from the first block of training trials (Figure 3)—the only block where all participants took part—were analyzed in a 2 (Melody Familiarity) x 2 (Picture Familiarity) between-subjects ANOVA. Effects of Melody Familiarity ($F(1,60) = 4.13$, $p = .046$) and Picture Familiarity ($F(1,60) = 4.05$, $p = .049$) were qualified by an interaction ($F(1,60) = 12.95$, $p = .0006$). T-tests suggested that for familiar pictures, there was a strong effect of melody familiarity ($t(30) = 4.04$, $p = .0003$), while for novel pictures, there was no effect of melody familiarity ($t(30) = 1.09$, $p = .28$).

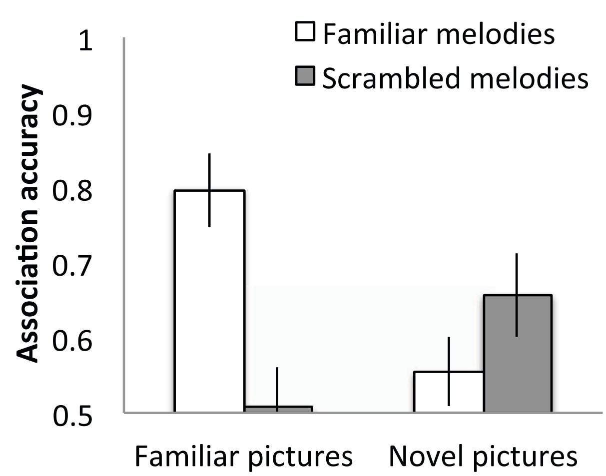


Figure 3. Experiment 1, accuracy in first block of training trials (± standard errors).

These results suggest that melody familiarity has strong effects on melody encoding, but only when the association is already somewhat familiar. This raises questions as to whether the familiarity advantage is at the level of perceptual representation, at the level of learned associations (with verbal materials or visual associates), or both. Therefore, the next experiment tested whether children perform better on familiar melodies in a task that more directly assesses perceptual processing: a same-different discrimination task. If children have more robust perceptual representations for familiar than for unfamiliar (scrambled) melodies, then discrimination performance should be better for changes from one familiar melody to another than for changes from one scrambled melody to another.

Because previous studies have suggested that children are more sensitive to timbre than to melodic contour for brief tone sequences (Creel, 2014, 2016), the next experiment also included timbre discrimination trials at two levels of timbre similarity.

IV. Experiment 2: Discrimination

A. Method

1) *Participants.* Sixty-four preschool-aged children (ages 4-5 years) who had not taken part in Experiment 1 participated. An additional 18 children took part but were replaced because of: failure to meet training criterion (13), computer error (3), failure to understand instructions (1), inattention (1).

2) *Stimuli.* Mary, Twinkle, Mary-scrambled and Twinkle-scrambled were heard by all participants. Participants also heard different-timbre trials where the same melody was played but the timbre changed. For half of participants, the two timbres were fairly similar (bassoon and alto saxophone); for the other half, the timbres were more distinct (vibraphone and muted trumpet). Timbre distinctiveness estimates were drawn from Iverson and Krumhansl (1993).

3) Procedure. Children received training trials on which highly-distinct melodies were used (rising vs. falling, harp vs. tuba, 1.5 octave difference in pitch range), including four “same” trials and four “different” trials in each block. If they did not achieve at least 7/8 correct, the training block repeated, up to 5 total training blocks. Once criterion was achieved, they continued to the test phase. The test phase presented all trials in a random order. Test stimuli included 8 melody-change trials (half familiar, half scrambled), 8 timbre-change trials (half using familiar melodies, half scrambled), and 16 same trials. An additional 8 trials presented the training stimuli (half same, half different) to assess continued task adherence.

B. Results and Discussion

Accuracy was analyzed in a 2 (Melody Familiarity) x 2 (Change Type) x 2 (Timbre Distance) ANOVA, with all but the last factor within-subjects. There was an overall effect of Melody Familiarity ($F(1,62) = 13.01$, $p = .0006$), with higher accuracy on trials using familiar melodies. There was also an effect of Change Type ($F(1,62) = 18.17$, $p < .0001$) such that overall, timbre changes were more accurately detected. However, Change Type interacted with Timbre Distance ($F(1,62) = 16.84$, $p = .0001$), which indicated that for similar timbre participants, there were no accuracy differences between melody-change and timbre-change trials ($t(31) = 0.11$, $p = .92$), while for the large timbre difference, the timbre-change trials were more accurate than the melody-change trials ($t(31) = 6.31$, $p < .0001$).

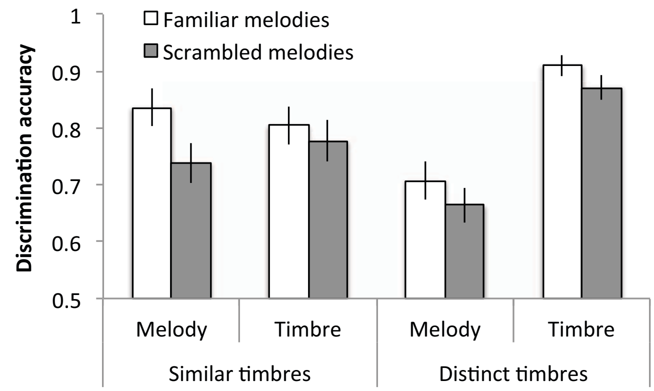


Figure 4. Experiment 2, discrimination accuracy, with standard errors.

Children appear to be better at distinguishing familiar melodies from each other than unfamiliar melodies, suggesting that they have more robust representations of familiar melodies. Interestingly, a large timbre difference still seems more salient than a difference between melodic contours even when those melodic contours are familiar.

V. General Discussion

The current study asked how children represent musical events: in terms of more abstract structures (musical scales), or in terms of specific memories? Results suggest that children represent specific memories: they are better at forming associations with familiar melodies than unfamiliar melodies (Experiment 1) and are better at distinguishing familiar melodies than unfamiliar melodies (Experiment 2). This facilitated processing of familiar melodies suggests that children’s musical representations may be fairly melody-specific.

This is not to suggest that there is no role for more abstract structures (or something like abstract structures) in musical representations. What it does suggest is that things like scale structures may arise from the statistics of one’s musical input (see, e.g., Pearce & Wiggins, 2012), perhaps over a long time period. It may be that only after a fairly long time period of musical enculturation are listeners able to represent unfamiliar melodies using scale structure—or perhaps that after such a long period of enculturation, the listener has such a large “mental songbook” that any new melody can be represented in terms of already-familiar songs.

Some limitations remain in interpreting these studies. For one thing, it is not clear why children would be less adept at mapping familiar melodies to novel objects than to familiar, melody-related objects. Strictly speaking, one would assume that a specific auditory representation would facilitate audiovisual association regardless of the familiarity of the picture. It may be that *both* picture familiarity and melody familiarity are needed. Alternatively, it may be that asking children to associate known music, which already has some semantic associations, with unfamiliar and seemingly unrelated objects makes the learning task especially difficult. One might wonder if pre-familiarizing children with melodies would get around this problem, but at least one previous attempt suggests that pre-familiarization in a lab setting is insufficient to generate enough familiarity to facilitate mapping.

Another limitation concerns the role of familiarity in melody discrimination. Familiarity effects were present, but not large. Why would effects be smaller for a perceptual discrimination task than an association task? There are many potential explanations, but one possibility is that effects are smallest when one can retain a melody in working memory. Once the listener is required to maintain a representation in a longer-term store—as would be required for association learning—weaker representations of unfamiliar melodies would be more apparent.

Nonetheless, results largely support the specific learning hypothesis. Models of musical development and musical representation may need to account for greater specificity in musical memory than is currently acknowledged.

ACKNOWLEDGMENT

Thanks to Emilie Seubert, Nicolle Paullada, and Elizabeth Groves for running child participants, and to the children, parents, and preschools who made this research possible.

REFERENCES

Corrigall, K. A., & Trainor, L. J. (2010). Musical enculturation in preschool children: acquisition of key and harmonic knowledge. *Music Perception*, *28*(2), 195–200.

Creel, S. C. (2014). Tipping the scales: Auditory cue weighting changes over development. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(3), 1146–1160. doi:10.1037/a0036057

Creel, S. C. (2016). Ups and downs in auditory development: Preschoolers' sensitivity to pitch contour and timbre. *Cognitive Science*, *14*, 373–303. doi:10.1111/cogs.12237

Dowling, W. J. (1978). Scale and contour: two components of a theory of memory for melodies. *Psychological Review*, *85*(4), 341–354.

Feldman, N. H., Myers, E. B., White, K. S., Griffiths, T. L., & Morgan, J. L. (2013). Word-level information influences phonetic learning in adults and infants. *Cognition*, *127*(3), 427–38. doi:10.1016/j.cognition.2013.02.007

Fennell, C. T., & Werker, J. F. (2003). Early word learners' ability to access phonetic detail in well-known words. *Language and Speech*, *46*(Pt 2-3), 245–64. doi:10.1177/00238309030460020901

Goldinger, S. D. (1998). Echoes of echoes?: An episodic theory of lexical access. *Psychological Review*, *105*(2), 251–279.

Hannon, E. E., & Trehub, S. E. (2005). Tuning in to musical rhythms: Infants learn more readily than adults. *Proceedings of the National Academy of Sciences*, *102*(35), 12639–12643.

Hannon, E. E., & Trehub, S. E. (2005). Metrical categories in infancy and adulthood. *Psychological Science*, *16*(1), 48–55.

Iverson, P., & Krumhansl, C. L. (1993). Isolating the dynamic attributes of musical timbre. *Journal of the Acoustical Society of America*, *94*(5), 2595–603. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8270737

Pearce, M. T., & Wiggins, G. A. (2012). Auditory expectation: the information dynamics of music perception and cognition. *Topics in Cognitive Science*, *4*, 625–652. doi:10.1111/j.1756-8765.2012.01214.x

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, *388*, 381–382.

Swingley, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*, *13*(5), 480–484. doi:10.1111/1467-9280.00485

Trainor, L. J., & Trehub, S. E. (1994). Key membership and implied harmony in Western tonal music: developmental perspectives. *Perception & Psychophysics*, *56*(2), 125–32. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7971113

Trehub, S. E., Bull, D., & Thorpe, L. A. (1984). Infants' perception of melodies: The role of melodic contour. *Child Development*, *55*(3), 821–830.

Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy*, *3*(1), 1–30. doi:10.1207/15250000252828226

Effects of dark and bright timbral instructions on acoustical and perceptual measures of trumpet performance

D. Gregory Springer[1] and Amanda L. Schlegel[2]
[1]School of Music, University of South Carolina, Columbia, SC, USA
[2]School of Music, University of Southern Mississippi, Hattiesburg, MS, USA

Prior research has indicated perceptual interactions among pitch, timbre, and loudness. For example, listeners often judge equal-frequency tones as different in pitch when their timbre varies, and listeners' loudness judgments are less accurate when comparison tones have different frequencies. The present study was designed to investigate how these perceptual interactions might be made manifest in a performance task. Specifically, we investigated the potential effects of timbral instructions (e.g., "Play with a dark sound") on the cent deviation, intensity, and tone quality of sustained tones. Collegiate trumpeters performed repetitions of two sustained tones (written C4 and C5) in three conditions: neutral instruction ("Perform this note"), dark-timbre instruction ("Perform this note with a dark sound"), and bright-timbre instruction ("Perform this note with a bright sound"). Four contrasting presentation orders were used. Dependent variables included two acoustic measures (cent deviation from equal-tempered standard frequency [expressed in absolute value] and intensity [expressed as dB]) and one perceptual measure (expert tone quality rating). Split-plot ANOVAs were conducted to examine the effects of instruction (neutral, dark, or bright), tone (C4 or C5), and order on each of these dependent variables. Preliminary results from a pilot study indicate that timbral instructions demonstrated a significant main effect on performed cent deviation, $F(2, 4) = 7.62$, $p = .04$, $\eta^2_p = .79$, and post-hoc comparisons indicate significantly higher deviation in response to the dark-timbre instruction than the neutral instruction ($p = .04$). No other main effects or interactions were observed in the pilot study. Data collection for the primary study (expected $N = 30$) is ongoing and will be complete by May 1. Results will provide useful implications for music pedagogy and further conclusions regarding perceptual interactions among pitch, timbre, and loudness.

Listeners rapidly adapt to musical timbre

Elise A. Piazza[1,2,3], Frédéric E. Theunissen[4], David Wessel[5,6], and David Whitney[4]
[1]Princeton Neuroscience Institute, Princeton University, Princeton, USA
[2]Helen Wills Neuroscience Institute, University of California, Berkeley, USA
[3]Vision Science Graduate Group, University of California, Berkeley, USA
[4]Department of Psychology, University of California, Berkeley, USA
[5]Department of Music, University of California, Berkeley, USA
[6]Center for New Music and Audio Technologies, University of California, Berkeley, USA

Timbre allows us to quickly identify threatening animal growls, pick out a friend's nasal voice among a crowd, and enjoy a rich variety of musical textures, from the bright buzz of a muted trumpet to the crisp attack of a snare drum. Often defined as tone color, or the unique quality of asound, timbre cannot be attributed simply to pitch, intensity, duration, or location. However, despite its critical contributions to our everyday auditory experiences, it is unclear whether timbre truly represents a meaningful auditory configuration rather than an arbitrary collection of low-level features: if so, listeners should adapt to timbre, and this adaptation should generalize across changes in lower-level features (e.g., pitch). Although adaptation to simple auditory features (e.g., frequency, level) is well studied, it is unknown whether or how adaptation to the more complex, holistic structure of natural sounds might influence perception. Here, we used a range of complex sounds equated in basic auditory features to investigate timbre adaptation. In each trial, participants adapted to one of two sounds (e.g., clarinet and oboe) that formed the endpoints of a morphed continuum. After repeated exposure to one of the two adapter sounds, participants judged the identity of the various morphed sounds. We found consistent negative aftereffects resulting from timbre adaptation, such that adapting to sound A significantly altered perception of a neutral morph between A and B, making it sound more like B. Importantly, these aftereffects were robust and invariant to changes in pitch, suggesting that they stem from adaptation to the global, spectro-temporal configuration of complex sounds and can survive the lower-level feature shifts that naturally vary (largely independently of timbre) in the environment. This adaptation, which is likely exploited by composers, could serve to enhance our sensitivity to novel or rare auditory objects, such as a new soloist in an orchestra.

All about that bass: Timbre and groove perception in synthesized bass lines

Ivan Tan[1], Ethan Lustig[1]
[1] Department of Music Theory, Eastman School of Music, University of Rochester, Rochester, NY, USA

Groove is defined as wanting to move the body to music. Most groove research has focused on temporal features like microtiming and syncopation, though previous studies have shown correlations between activity in certain frequency bands and increased movement. This study further examines the relationship between timbre and groove. Four brief songs consisting of drum samples and a bass synthesizer were composed in an electronic dance music style and presented to 102 participants on headphones. Each song had four audio filter conditions (no filter, low-pass, high-pass, band-pass) applied only to the synthesizer. Participants heard each stimulus three times, and rated both groove and liking on a five-point scale.

For both groove and liking, a main effect was found for filter—participants gave higher ratings to the filter conditions that preserved low-frequency energy (low-pass and no filter), and lower ratings to the filter conditions that removed low-frequency energy (high-pass and band-pass). No song-filter interaction was found, and no repetition effect was found.

The results suggest—not surprisingly—that people groove more to, and like more, the filter conditions that preserve low-frequency energy. The lack of song-filter interaction demonstrates that within the same style, timbre can determine groove and liking, even across different melodic and rhythmic contexts. Our results are consistent with other recent empirical groove research on timbre, and we consider recent physiological explanations for these findings. We also analyze our data for gender differences in filter ratings, and consider cultural and evolutionary explanations, as some research suggests that males prefer exaggerated bass. Finally, we consider the effect of dance experience on filter ratings, based on questionnaire data.

Mobile Experience Sampling of Music Listening using the MuPsych App

Will Randall[1], Suvi Saarikallio[1]
[1]Department of Music, University of Jyväskylä, Jyväskylä, Finland

Personal music listening on mobile phones is rapidly growing as a central component of everyday musical engagement. This portable and flexible style of listening allows for the immediate selection of music to fulfil emotional needs, presenting it as a powerful resource for emotion regulation. The experience sampling method (ESM) is ideal for observing personal music listening, as it assesses current subjective experience during natural everyday music episodes. The MuPsych app utilises mobile ESM, and was developed for the ecologically valid and real-time measurement of emotional responses to music listening on mobile devices. This workshop will demonstrate to researchers how to design and implement their own experience sampling studies using the existing MuPsych mobile app. An online interface will allow researchers to easily customise their own study, including their own questions presented during listening episodes, along with psychological surveys. This presents an opportunity for a broad range of music researchers to tie their work to everyday musical experiences, and benefit from the real-time measurement and ecological validity of the mobile ESM.

BeatHealth: Bridging rhythm and technology to improve health and wellness

Simone Dalla Bella[1], Ramesh Balasubramaniam[2]
[1]EuroMov Laboratory, Montpellier University, Montpellier, France
[2]Cognitive & Information Sciences, University of California - Merced, Merced, California, USA

There is a tight link between the rhythm of external information (e.g., auditory) and movement. It can be seen when we spontaneously or deliberately move on the beat of music (e.g., in dance), or when we particularly enjoy performing sport activities (e.g., running or cycling) together with music. The propensity to match movements to rhythm is natural, develops precociously and is likely hard-wired in humans. The BeatHealth project exploits this compelling link between rhythm and movement for boosting motor performance and enhancing health and wellness using a personalized approach. This goal is achieved by creating an intelligent technological architecture for delivering rhythmic auditory stimuli using state-of-the art mobile and sensor technology. Optimal stimulation is based on experiments performed in the lab and in more ecological conditions. The beneficial effects of BeatHealth are evaluated in two populations: patients with Parkinson's disease (walking) and in healthy citizens of various ages with moderate physical activity (running). The main findings and technological achievements of this project, still ongoing, will be presented in this Symposium.

Optimization of rhythmic auditory cueing for gait rehabilitation in Parkinson's disease

Simone Dalla Bella[1,2,3], Dobromir Dotov[1], Sophie Bayard[4], Valérie Cochen de Cock[5,6], Christian Geny[5], Emilie Guettard[6], Petra Ihalainen[1], and Benoît Bardy[1,2]
[1]EuroMov Laboratory, Montpellier University, Montpellier, France
[2]Institut Universitaire de France (IUF), Paris, France
[3]International Laboratory for Brain, Music, and Sound Research (BRAMS), Montreal, Canada
[4]Department of Psychology, Paul Valery University, Montpellier, France
[5]Centre Hospitalier Universitaire de Montpellier, Montpellier, France
[6] Beau Soleil Clinic, Montpellier, France

Gait dysfunctions in Parkinson's disease (PD) can be partly relieved by rhythmic auditory cueing (i.e., asking patients to walk with a rhythmic auditory stimulus such as a simple metronome or music). The beneficial effect on gait is visible immediately in terms of increased speed and stride length. In standard rhythmic auditory cueing, the interval between sounds or musical beats is fixed. However, this may not be an optimal stimulation condition, as it fails to take into account that motor activity is intrinsically variable and tends to be correlated in time. In addition, patients show typical variability (accelerations and decelerations) during walking which may lead them to progressively desynchronize with the beat. Stimulation with embedded biological variability and capable to adapt in real time to patients' variable step times could be a valuable alternative to optimize rhythmic auditory cueing. These possibilities have been examined in the BeatHealth project, by asking a group of PD patients (n = 40) and matched controls (n = 40) to walk with various rhythmic auditory stimuli (a metronome, music, and amplitude-modulated noise) under different conditions. In some conditions, inter-beat variability has been manipulated, by embedding biological vs. random variability in the stimulation. Another manipulation was to adapt in real time the stimulus beats to participants' step times using dedicated algorithms. Performance was quantified by way of the typical spatio-temporal gait parameters and additional measures akin to complex systems (e.g., long-range correlations). The results indicated that biologically-variable stimulation which is adapted to patients' steps is the most efficient for improving gait performance. These findings provide useful guidelines for optimizing music-driven rehabilitation of gait in PD.

When music (de-)synchronizes biology during running

Benoît Bardy[1,2], Loïc Damm[1], Fabien Blondel[1], Petra Ihalainen[1], Déborah Varoqui[1] and Simone Dalla Bella[1,2]
[1]EuroMov Laboratory, Montpellier University, Montpellier, France
[2]Institut Universitaire de France, Paris, France

Although the inclination to run in music seems to transcend places, generations, and cultures, it is unclear where this ergogenic power of music comes from. During running, it can be merely due to music's tendency to elicit pleasure, increase motivation or drive attention away from athletes' feeling of fatigue. Alternatively, music may foster efficient coupling of biological rhythms that are key contributors to performance, such as locomotion and respiration. In this talk, we present the main concepts underlying rhythmical entrainment in biological systems, and their potential consequences in terms of movement stability, overall performance and wellness. We present the results of several experiments performed in young adults running on a treadmill or outside the lab. These experiments exploit our intelligent running architecture — BeatRun — able to deliver embodied, flexible, and efficient rhythmical stimulation. Our results show the often beneficial but sometimes detrimental effect of running in music, the entrainment asymmetry between locomotor and respiratory systems, and large individual differences. These results are discussed in terms of the dynamics of flexible coupling between running and breathing and rhythmical entrainment. Consequences for sport performance, wellness, efficiency in daily work, and the development of new music technologies are envisaged.

The empowering effect of being locked to the beat of music

Marc Leman[1], Edith Van Dyck[1], Bart Moens[1] and Jeska Buhman[1]
[1] IPEM – Department of Musicology, Ghent University, Ghent, Belgium

A regular rhythm in music is a strong driver for establishing a synchronized human rhythm. Typically human rhythm tends to go along with the musical rhythm in such a way that a salient feature of the human rhythm matches the timing of a salient feature of the musical rhythm. But what are the underlying mechanisms behind music-movement synchronization, and what are the generated effects? The goal of this talk is to introduce and offer different views on the study of the underlying mechanisms, and to show how being locked to the beat of the music can pave the way for an overall empowerment effect. We base our ideas on a technology that allows us to manipulate the entrainment strength towards phase-locked synchronization. The results of our findings are relevant for assistive synchronization technologies such as a smart walkman for patients with Parkinson's disease, or a running assistent for healthy runners developed in BeatHealth.

BeatHealth: a platform for synchronizing music to movement in the real world

Rudi Villing[1], Joseph Timoney[2], Victor Lazzarini[3], Sean O'Leary[1], Dawid Czesak[1], and Tomas Ward[1]
[1]Department of Electronic Engineering, Maynooth University, Maynooth, Ireland
[2]Department of Computer Science, Maynooth University, Maynooth, Ireland
[3]Department of Music, Maynooth University, Maynooth, Ireland

The BeatHealth project delivers a music driven application designed to improve the wellness of healthy runners and people with Parkinson's disease by suitably modulating the coupling of music to their movement. Specifically, musical stimulation is adapted in real time to synchronize with human gait so as to motivate the user and reinforce, stabilize, or modify their gait characteristics. This research focused on the development of a mobile platform, comprising a mobile phone application and low-cost wearable wireless sensors. This mobile platform conveniently translates the BeatHealth laboratory findings and algorithms to an embodiment suitable for real world use. The development process demanded an interdisciplinary approach which integrated the fields of movement, music, and technology. Sensor technology was developed to accurately track the kinematic activity which formed the input to the musical stimulation process running on the mobile phone. This process selects suitable audio tracks from a playlist and dynamically alters their playback rate in real time so that the rhythm of the movement and the music are synchronized. Multiple sources of delay and variability in the technology were measured and compensated for to ensure that the desired synchronization could be achieved. The system also required tuning so that it responded appropriately to natural movement variability and intentional movement changes. The mobile system has been validated in a variety of settings using laboratory equipment. Ultimately, the BeatHealth mobile platform provides a synchronization capability that has not been available to date for music and movement applications. Moreover, the technical architecture of the platform can be adapted to synchronize with essentially any repetitive movement or rhythmic pulse which can be described as an oscillator. This gives the platform potential applicability in a wide range of studies examining the complex relationship between music and movement.

Leveraging large-scale industrial data to study musical engagement

Blair Kaneshiro[1,2] and Douglas Eck[3]
[1]Shazam Entertainment Ltd., Redwood City, CA, USA
[2]Center for Computer Research in Music and Acoustics, Stanford University, Stanford, CA, USA
[3]Google Research, Mountain View, CA, USA

In recent years, mobile, online, and social-media services have yielded an unprecedented volume of user data reflecting various forms of musical engagement. While these data typically lack the underlying experimental rigor of controlled laboratory studies, they surpass traditional experimental data in size, scope, and ecological validity. Analytical insights drawn from such data thus promise to provide significant contributions to our understanding of numerous musical behaviors. Embodying the Bay Area's unique integration of academia and industry, this symposium aims to introduce the ICMPC community to music cognition research conducted in an industrial setting. We convene researchers from four well-known music technology companies, who will each explore a different aspect of musical engagement: Discovery (Shazam), performance (Smule), education (Yousician), and recommendation (Pandora). The symposium will comprise back-to-back presentations followed by a short panel discussion and audience Q&A facilitated by the discussant (Google). All symposium participants have research experience in both academic and industrial settings, and are excited to engage in a discussion of research goals, practices, and practicalities bridging academia and industry. It is hoped that this symposium will open a dialogue between academia and industry regarding potential benefits and challenges of incorporating industrial data into music perception and cognition research. The symposium will be relevant to conference delegates performing research in any of the above topics, as well as anyone interested in using industrial data to broaden the scope of research in musical practices and behaviors. Attendees will come away with a practical understanding of requisite tools and technical skills needed to perform music cognition research at scale, as well as perspective on the impact of data size on hypothesis formulation, experimentation, analysis, and interpretation of results.

Musical correlates of large-scale song discovery: A Shazam study

Blair Kaneshiro[1,2] and Casey W. Baker[1]
[1]Shazam Entertainment Ltd., Redwood City, CA, USA
[2]Center for Computer Research in Music and Acoustics, Stanford University, Stanford, CA, USA

When a listener deliberately seeks out information about a piece of music, a number of cognitive, contextual, and musical factors are implicated – including the listener's musical tastes, the listening setting, and features of the musical performance. Such drivers of music discovery can now be studied on a massive scale by means of user data from Shazam, a widely used audio identification service that returns the identity of an audio excerpt – usually a song – in response to a user-initiated query. Shazam's user data, comprising over 600 million queries per month from over 100 million monthly users across a catalog of over 30 million tracks, have been shown to predict hits, reflect geographical and temporal listening contexts, and identify communities of listeners. We begin this talk with an overview of the Shazam use case and its user data attributes. Next, we present findings from an exploratory study utilizing Shazam query 'offsets', the millisecond-precision timestamps specifying when in a song a query was initiated. Here we wish to see whether there exists a relationship between the distribution of offsets over the course of a song and specific musical events. We aggregated over 70 million offsets corresponding to ten top hit songs of 2015. Taking into account the estimated delay between intent-to-query and actual query times, results thus far suggest that offsets are not uniformly distributed over the course of a song, but in fact cluster around specific musical events such as the initial onset of vocals and transitions between song parts. We consider the implications and potential confounds surrounding these findings in terms of music-listening contexts, distinguishing between music discovery and musical tastes, and practical considerations surrounding Shazam's functionality. We conclude with a discussion of potential uses of Shazam data to address central questions in the music perception and cognition field.

Measuring musical inhibition: A demographic analysis of sharing performances on social media

Jeffrey C. Smith[1,2]
[1]Smule Inc., San Francisco, CA, USA
[2]Center for Computer Research in Music and Acoustics, Stanford University, Stanford, CA, USA

The character and nuance of musical performances constitute a particular sense of excitement and pleasure in live performances, afford the opportunity for the audience to identify directly with the artist, and indeed comprise a fundamental aspect of western music. Performing artists strive to exude a sense of command by balancing confident predetermined interpretive gesture with a sense of spontaneity and freedom. However, for many, achieving this balance can be difficult and fraught with anxiety. Overcoming performance inhibition in musical performance has been a topic of study in psychology. In this talk we describe a novel framework for measuring performance inhibition using a large database of amateur, aspiring professional, and professional singers' performances. The performances are recorded using the Smule Sing! Karaoke Application for mobile phones. Users have the option to share their performances with a select group or the general public. Of the 12 million performances recorded daily, around one million are shared publicly. Virtually all such shared performances include geographical and demographic information about the performer, and often include video. The broad range of expertise, stylistic, and demographic factors in the data provides a unique opportunity to study performance practices with far greater breadth than studies of professional recording artists or pre-professional music students that are typically the source of most existing studies. We will attempt to draw observations on performance inhibition from aspects of the performance as well as the singer's decision to release the performance. We will elaborate on the data source, the process to extract the data, and the value as well as limitations of this corpus. We attempt to measure such inhibition in a relative sense considering demographics including age, gender, or regional tendencies. The data set and analytical tools used here are available to researchers by arrangement.

Learning to play a musical instrument: Individual differences and commonalities

Anssi Klapuri[1], Antti Hurmalainen[1] and Oguzcan Sehiralti[1]
[1]Yousician, Helsinki, Finland

We studied the learning process of students who practice to play a musical instrument in the self-study setting. The goal was to discover individual differences and commonalities in the students' practice habits and learning efficiency. Data collection took place via Yousician, a music education platform with over five million users. In the application, users are prompted to perform an exercise and the acoustic signal from the instrument is then used to evaluate whether the correct notes were played at the right times, and to deliver appropriate real-time feedback. Students tend to develop varying practice habits, even in the restricted setting where the syllabus is pre-determined: Some prefer to perfect each exercise before moving on to the next one, whereas others push forward faster. Some revisit previously learned material from time to time, whereas others do not – just to mention a few examples. We carried out statistical analysis on a random set of 25,000 subscribed users, in order to study the correlation between certain practice habits and the resulting learning efficiency. Here learning efficiency was defined as the cumulative practice time required to pass skill tests at varying levels of difficulty. We also studied how pitch and timing inaccuracies in users' performances correlate with the music content, in order to identify factors that make a song difficult to perform. The use of such data for optimizing computer-assisted learning and the syllabus contents will be discussed.

Personalized music recommendation with Pandora

Oscar Celma[1]
[1]Pandora Media Inc., Oakland, CA, USA

Pandora began with The Music Genome Project, the most sophisticated taxonomy of musicological data ever collected and an extremely effective content-based approach to music recommendation. Its foundation is based on human music cognition, and how an expert describes and perceives the complex world of a music piece. But what happens when you have a decade of additional data points, given off by over 250 million registered users who have created 8+ billion personalized radio stations and given 60+ billion thumbs? Unlike traditional recommender systems such as Netflix or Amazon, which recommend a single item or static set, Pandora provides an evolving set of sequential items, and needs to react in just milliseconds when a user is unhappy with a proposed song. Furthermore, several factors (musicological, social, geographical, generational) play a critical role in deciding what music to play to a user, and these factors vary dramatically for each listener. This session will look at how Pandora's interdisciplinary team goes about making sense of these massive data sets to successfully make large-scale music recommendations. Following this session the audience will have an in-depth understanding of how Pandora uses Big Data science to determine the perfect balance of familiarity, discovery, repetition and relevance for each individual listener, measures and evaluates user satisfaction, and how our online and offline architecture stack plays a critical role in our success. I also present a dynamic ensemble learning system that combines curational data and machine learning models to provide a truly personalized experience. This approach allows us to switch from a lean-back experience (exploitation) to a more exploration mode to discover new music tailored specifically to users' individual tastes. I will also discuss how Pandora, a data-driven company, makes informed decisions about features that are added to the core product based on results of extensive online A/B testing.

The role of updating executive function in sight reading performance across development

Laura Herrero,*[1] Nuria Carriedo,#[2] Antonio Corral#[2], Pedro R. Montoro#[2]

*Education Dept., Camilo José Cela University, Spain

#Development and Education Psychology Dept., National Distance Education University (UNED), Spain

[1]lherrero@ucjc.edu, [2]ncarriedo@psi.uned.es, [2]acorral@psi.uned.es, [2]prmontoro@psi.uned.es

ABSTRACT

The main aims of our study were: (1) to explore the relationship between updating executive function sub-processes—maintenance and inhibition— and Sight Reading (SR) performance across development and different levels of expertise, and (2) to analyze the existence of individual differences between good and poor sight readers that could be associated with updating executive function subprocesses. Participants were 131 students ranging from 10 years old children who had received 2 years of music training in string or wind instruments, to young adults who had received training for more than 12 years. SR was measured in terms of rhythm accuracy, using two different pieces (binary and ternary time bar) by level of expertise extracted from "Sound and sight series" (Trinity College London). Updating executive function was measured by the updating task devised by De Beni and Palladino (2004), which allowed to differentiate between retrieval/transformation and substitution processes: maintenance, suppression of information in WM, and proactive interference. The results showed that rhythm accuracy in SR was significantly related both to maintenance and suppression of information in WM regardless the age and expertise level, whereas there were no significant correlations between rhythmic accuracy in SR and proactive interference. The results also showed that good sight readers outperformed poor sight readers both in maintenance and suppression of information in WM.

I. INTRODUCTION

Music sight reading (SR), defined as *"the ability to play music from a printed score or part for the first time without benefit of practice"* (Wolf, 1976, p.143), has been considered as one of the five basic abilities for all the musicians, together with rehearsed repertory performance, performance from memory, performance by ear, and improvisation (McPherson & Thompson, 1998). Thus, contrary to rehearsed music performance, SR performance would be considered as a novel task without these practice benefit.

Although SR expertise has been revealed as the most significant predictor of SR performance (Kopiez & Lee, 2006, Kopiez & Lee, 2008; Kopiez, Weihs, Ligges, & Lee, 2006; Lehmann & Ericsson, 1996; Lehmann & McArthur, 2002; Meinz & Hambrick, 2010), from a cognitive perspective, SR has been described as a complex transcription task that simultaneously involves reading of new non-rehearsed material and performance (Sloboda, 1982; Thompson, 1987; Waters, Townsend, & Underwood, 1998; Waters, Underwood, & Findlay, 1997). Specifically, Waters et al. (1998) suggested that SR achievement could be determined by efficiency in processing the incoming information during the translation into movements of music printed in the score. Thus, performance at sight could be considered as a complex dual task because it would imply cognitive resources linked to reading, decoding and understanding of musical structure—including rhythm and pitch—, as well as resources linked to performance—including fine motor control and musicality—, all in real time.

The study of acquisition of SR skills in novices and experts has allowed to determine the cognitive abilities underlying SR. In this context, some authors such as Drake and Palmer (2000) suggested that skilled sight readers, in comparison to novices, could have improved memory associations between musical fragments or sequences, which in turn would increase their ability to simultaneously plan and execute. Lehmann and Kopiez (2009) also described memory processes linked to SR trough the perceptual eye-hand span, considered as "*the amount of information stored temporarily from a particular sequence of fixations or during a certain timespan"* (p.346). These authors concluded that eye movement were reduced as an inversely function of the level of expertise, and that better sight readers could anticipate the execution and do it more fluent. Given that both planning and anticipation abilities in SR could imply high demands on working memory (WM) to simultaneously maintain and process the information during motor performance (Kopiez & Lee, 2008), and due to the temporary nature of the task, it could be necessary the continuous updating of information in WM.

Updating information in WM has been considered, together with shifting and inhibition, as one of the executive functions described in adults (Miyake et al., 2000) as well as in children and adolescents (Huizinga, Dolan, & van der Molen, 2006; Lehto, Juujärvi, Kooistra, & Pulkkinen, 2003; Xu, Han, Sabbagh, Wang, Ren, & Li, 2013). Updating was initially defined by Morris and Jones (1990), as the mechanism responsible of the replacement of WM content that is no longer relevant for the ongoing task for new novel material, and the adjustment of the remained material to the incoming new. Specifically, updating information in WM has been considered a significant predictor of individual differences in fluid intelligence (Chen & Li, 2007) and in reasoning (Barrouillet, 1996; Barrouillet & Lecas, 1999; Kyllonen & Christal, 1990), and also in other high cognitive abilities such as mathematical achievement (Bull & Lee, 2014; Iglesias-Sarmiento, Carriedo & Rodríguez, 2014; Passolunghi & Pazzaglia, 2005), or reading comprehension (Daneman & Carpenter, 1980; Iglesias-Sarmiento et al., 2014; Palladino, Cornoldi, De Beni, & Pazzaglia, 2002).

From different perspectives, updating executive function has been highly related to WM capacity (Belacchi, Carretti, & Cornoldi, 2010; Miyake et al., 2000; St Clair-Thompson & Gathercole, 2006). In this context, Ecker, Lewandowsky, Oberauer, and Chee (2010) analyzed this relationship using structural equation models (SEM). They concluded that WM capacity and updating executive function could be distinguishable in terms of their underlined processes. Specifically, they described three processes underlying updating executive function: a) retrieval of the relevant information from long term memory (LTM); b) transformation of information in WM; and c) substitution of information in WM, which inhibits and replaces information that is no longer relevant. Both retrieval and transformation sub-processes could represent a common source of variance in WM capacity and updating, whereas substitution could be specific of the updating function, and could allow the differentiation between updating and WM capacity (see also Ecker, Lewandowsky & Oberauer, 2014). Taking into account the cognitive requirements of SR music, and the involvement of updating executive function in high level cognition, it could be supposed that updating could be in part responsible of the differences in this complex task. Moreover, we also consider that the underlying processes of retrieval/transformation and substitution described above could contribute differentially to SR performance at both efficiency and developmental levels: the processes of retrieval and transformation could be necessary to maintain in WM the information printed in the score and to select the adequate motor responses from LTM; on the other hand, the process of substitution could be necessary to supress no longer relevant information from WM, replace it for the new incoming information from the score, and inhibit inadequate motor responses from LTM..

To our knowledge, it has not yet been explored the role of updating executive function underlining processes in SR performance. Moreover, the vast majority of studies considering cognitive processes in SR were carried out with adult and expert pianist participants. For this reason, we established two main goals: first, to explore the relationship between updating executive function sub-processes and SR rhythmic accuracy across development and expertise levels; and second, to analyze the existence of differences between good and poor sight readers, in terms of rhythmic accuracy, that could be associated with updating executive function sub-processes.

We first hypothesized that, regardless of the effect of age and expertise level, the processes of retrieval/transformation and substitution underlying updating executive function would be significantly related to SR rhythmic accuracy. Second, we also predicted that good sight readers would outperform poor sight readers in the updating task, which will be reflected both in the processes of retrieval/transformation and substitution.

II. METHOD

A. Participants

The sample for this study consisted of 131 students of different ages and level of expertise: 37 children aged 10-11 (*M* = 10.9, *SD* = .44; 26 girls, 11 boys), that have been receiving music lessons since at least 2 years (*M* = 2.9, *SD* = .32) of melodic instruments (10 violin, 3 viola, 3 cello, 3 double bass, 9 flute, 1 oboe, 2 clarinet, 3 saxophone, 1 trumpet, 2 trombone) formed the novices-low expertise group; 31 pre-adolescents aged 12-13 (*M* = 13.2, *SD* = .48 ; 25 girls, 6 boys), that have been receiving music lessons since at least 4 years (*M* = 4.6, *SD* = .44) of melodic instruments (11 violin, 4 viola, 3 cello, 1 double bass, 4 flute, 1 oboe, 4 clarinet, 2 saxophone, 1 trumpet) formed the novices-high expertise group; 32 adolescents aged 15-16 (*M* = 15.2, *SD* =.49 ; 21 girls, 11 boys), that have been receiving music lessons since at least 7 years (*M* = 6.8, *SD* = .28) of melodic instruments (12 violin, 1 viola, 4 cello, 2 flute, 4 oboe, 4 clarinet, 2 saxophone, 1 horn, 2 trumpet) formed de intermediate expertise group; and 31 young adults (*M* = 22.3, *SD* = 1.34; 14 women, 17 men), that have been receiving music lessons since at least 12 years (*M* = 12.9, *SD* = .46) of melodic instruments (2 violin, 1 viola, 2 cello, 8 flute, 4 oboe, 2 clarinet, 2 bassoon, 2, saxophone, 4 horn, 1 trumpet, 2 trombone, 1 tuba) formed the high expertise group.

B. Materials

Updating Task. We used an adaptation to the Spanish of the updating task (Carriedo, Corral, Montoro, Herrero & Rucián, 2016) developed by De Beni and Palladino, (2004). This updating task allows differentiating retrieval/transformation and substitution components of the updating process, posing variable demands on memory and suppression, which permits the study of the interaction of the two processes. Importantly, this task also provides different indexes for the process of inhibition/interference: suppression of information in WM, that is, the ability to inhibit in WM information that is irrelevant to the goals of the task, and proactive interference, understood as the capacity to inhibit the information stored in LTM that is no longer relevant for the ongoing task. This task was selected because it has been shown to be a reliable measure in the study of individual differences in various areas, such as arithmetic problem-solving (Iglesias-Sarmiento et al. 2014;Passolunghi & Pazzaglia, 2005), reading comprehension (Carretti, Borella, Cornoldi, & De Beni, 2009; Iglesias-Sarmiento et al. 2014), and in the description of cognitive changes associated with age (Carriedo et al., 2016; De Beni & Palladino, 2004; Lechuga, Moreno, Pelegrina, Gómez-Ariza, & Bajo, 2006).

The task had a total of 24 lists (20 experimental lists and 4 practice lists), each one containing 12 words. The words were names of objects, animals, or body parts of different sizes, and abstract common nouns. Each list included words to be recalled (relevant words), words to be discarded (irrelevant words), and filler words. The number of each kind of word in each list varied depending on the experimental condition. Thus, the number of relevant words in each list varied between 3 (low memory load) and 5 (high memory load). The number of irrelevant words varied between 2 (low suppression) and 5 (high suppression). Finally, the number of abstract filler words varied from 2 to 7. The composition of lists as a function of memory load and suppression can be seen in Table 1. The final 24 lists were distributed in 4 experimental conditions of 6 lists each. One list of each experimental condition was considered as a practice list. The total words to be recalled across all condition lists were 80 (practice lists were excluded), 25 in the case of each high load condition, and 15 in the case of each low load condition. A list example is: *árbol (tree), autobús (bus), piscina (pool), sofá (couch), cesta (basket), tema (matter), acto (act), flor (flower), dedo (finger), lápiz (pencil), oreja (ear), patata (potato).*

Dependent variable was the percentage of correct recall as an index of retrieval/transformation of information in WM; same list intrusions as an index of suppression of information in WM; and previous lists intrusions as an index of proactive interference.

Sight Reading Task. Two different pieces from “Sound at Sight Series” (Trinity College London) were selected by instrument specialists for each instrument and level. The selection criterion was the time bar, choosing one binary (A) and one ternary (B) time bar for all the participants. Taking into account the musical knowledge by level of expertise, we decided to select a 2/4 or 4/4 piece as a binary time bar for all the levels and instruments. Following the same criteria, we selected the ternary pieces according to both time bar and subdivision. Thus, we selected a 3/4 piece (ternary time bar and binary subdivision) for the novices-low expertise group, a 6/8 or 12/8 piece (binary time bar and ternary subdivision) for the novices-high expertise group, and a 6/8, 12/8, 3/8 or 9/8 (binary or ternary time bar and ternary subdivision) for the intermediate and expertise levels. We selected rhythmic accuracy as an index of SR performance taking into account that previous research has been determined that it is a reliable measure of SR execution (e.g. Drake & Palmer, 2000; Elliot, 1982; Gromko, 2004; Hayward & Gromko, 2009; Kopiez & Lee, 2008; McPherson, 1994; Mishra, 2014; Waters et al.,1997; Wurtz, Mueri, & Wiesendanger, 2009). Rhythmic accuracy was measured by the correct proportion between notes and its corresponding rests.

Table 1. Composition of the lists as a function of the experimental conditions.

Low load/low suppression (5 Lists)	Low load/high suppression (5 Lists)	High load/low suppression (5 Lists)	High load/high suppression (5 Lists)
3 Relevant items 2 Irrelevant items 7 Filler items	3 Relevant items 5 Irrelevant items 4 Filler items	5 Relevant items 2 Irrelevant items 5 Filler items	5 Relevant items 5 Irrelevant items 2 Filler items

C. Procedure

All participants were tested individually. The order of administration of the Updating Task and the SR Tests was counterbalanced. The order of the binary (A) and ternary (B) SR Tests was also counterbalanced. All participants were evaluated at their music schools.

In the updating task, twenty-four lists, each comprising twelve auditory words, were presented to participants. Within each list, the auditory stimuli lasted less than 1000 ms and were presented at a constant speed with a 2-s interval between words. Presentation order of the lists was randomized and, additionally, within each list, item presentation was also randomized. Randomization and time were controlled by E-Prime software, version 2.0 (Psychology Software Tools Inc; www.pst-net.com/eprime). Participants were instructed to carefully listen to the list and, when it finished, to recall the 3 or 5 smallest animals or objects of the list. At the beginning of each list, a text message was displayed on a computer screen to indicate the concrete number of smallest items to remember. Then, a beep preceded the first word of the list. At the end of each list of 12 items, a different beep and a big question mark on the screen asked the participants to recall the 3 or 5 smallest items of the current list by verbal response. To continue with the next list, participants had to press the space bar.

In the SR task, all participants performed two consecutive scores. They were instructed to carefully look at the first score (tempo, bar, key signature), and 30 seconds later they were instructed to play trying not to make breakdowns or repetitions. The same procedure was carried out for the second score. All the performances were audio-registered (SONY® IC Recorder ICD-UX71F) and scored by two independent expertise musician judges.

III. RESULTS

For the analysis of the results, we previously tested whether there were differences by age and level of expertise in binary and ternary SR rhythmic accuracy measures. Thus, two independent univariate ANOVAs were carried out. The results showed that there were no significant differences in both, binary (F (3, 127) = .533, p = .66, ηp^2 = .01) and ternary (F (3, 127) = .554, p = .65, ηp^2 = .01) SR tests by age and level of expertise. Thus, we grouped all the participants for subsequent analysis.

Next, we tested whether binary and ternary SR tests were significantly different. Although there was a highly significant correlation between them (r = .917**, p < .001), Student t, test showed that ternary test means were significantly lower than binary ones (t = 3.574, p < .001) for all the participants. Thus, we decided to analyze them separately.

In order to test our first hypothesis, that is, that the processes of retrieval/transformation and substitution underlying updating executive function would be significantly related to SR rhythmic accuracy, a correlational analysis was carried out. Pearson correlations are shown in Table 2.

Table 2. Results of Pearson´s correlations (*p<.05, **p<.01).

	Low Load		High Load	
	Low supression	High Supression	Low Supression	High supression
Panel A. Recall of critical words.				
Binary	.14	.10	**.25****	.10
Ternary	.12	.11	**.25****	.09
Panel B. Same-list intrusions.				
Binary	-.11	-.09	**-.23****	-.11
Ternary	-.02	-.06	-.13	-.09
Panel C. Previous-list intrusions.				
Binary	-.00	-.09	-.04	-.06
Ternary	-.02	-.14	-.04	-.05

As it shown, the only significant correlations were obtained between the recall of critical items in the high load and low suppression condition and SR rhythmic accuracy in binary and ternary SR rhythmic accuracy and recall of critical items in high load and low suppression condition. Moreover, binary SR rhythmic accuracy was correlated to same-list intrusions. Thus, both maintenance and and suppression of information in WM were significantly correlated to binary SR rhythmic accuracy, whereas only maintenance was significantly related to ternary SR rhythmic accuracy. The results did not show significant correlation between rhythmic performance and proactive interference.

Once analysed the correlations, and taking into account our second aim, that is, to determine the existence of differences between good and poor sight readers in terms of rhythmic accuracy that could be associated with updating sub-processes. Due that there were no significant differences in SR rhythmic accuracy by age, we decided to calculate the frequencies for both binary and ternary tests to select two groups of participants, that is, good and poor sight readers. In base of superior and inferior quartiles, two groups were formed: (1) Good sight readers (N = 38), including the participants located in superior quartile both in binary and ternary tests; (2) Poor sight readers (N = 38), including the participants located in inferior quartile both in binary and ternary tests. Table 3 shows the distribution by age and level of expertise in each group.

Table 3. Distribution of good and poor sight readers by age and level of expertise.

	Poor sight readers	Good sight readers
Novices low	11	10
Novices high	10	10
Intermediate	9	6
Expert	8	12
Total N	38	38

For all the dependent variables we conducted a mixed ANOVA 2 x (2 x 2), with SR group (good/poor) as a between-subject factor, and memory load (high and low) and level of suppression (high and low) as within-subject factors. One-tailed test results are shown, unless otherwise is stated. Figure 1 shows the mean proportion of correct answers and errors by group (good and poor sight readers).

1) Proportion of recall of critical items.

The results showed a significant effect of memory load(F (1, 74) = 200.605, p = .000, ηp^2 = .73), which means that the proportion of correct responses was lower in the high load (75%) condition than in the low load (92%) condition for all the participants. Importantly, the results also showed a significant effect of group (F (1, 74) = 5.043, p < .002, ηp^2 = .06), which means that the proportion of correct responses was significantly higher in good sight readers (86%) than in poor sight readers (81%).

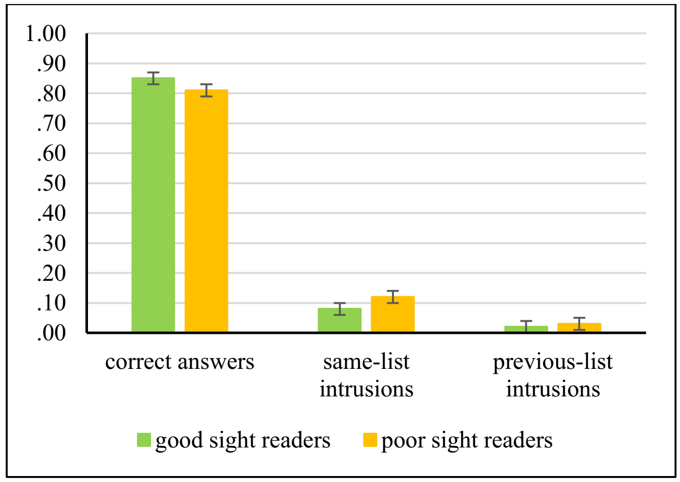


Figure 1. Mean proportion of correct responses and errors by group.

2) Proportion of same-list intrusions.

The ANOVAs showed significant main effects for memory load (F (1, 74) = 48.101, p = .000, ηp^2 = .39), and for level of suppression (F (1, 74) = 34.250, p =.000, ηp^2 = .32). The proportion of same-list intrusions was higher in the high load (14%) condition than in the low load (5%) condition. It was also higher in the high suppression (13%) condition than in the low suppression (7%) condition for all the participants. More relevant for our study, the results also showed a significant interaction memory load x group: F (1, 74) = 2.89, p < .05, ηp^2 = .04, which means that the effect of memory load in same-list intrusions was lower in good sight readers (low load 4% vs high load 12%) than in poor sight readers (low load 6% vs high load 17%). Figure 2 shows the proportion of same-list intrusions by group and memory load conditions.

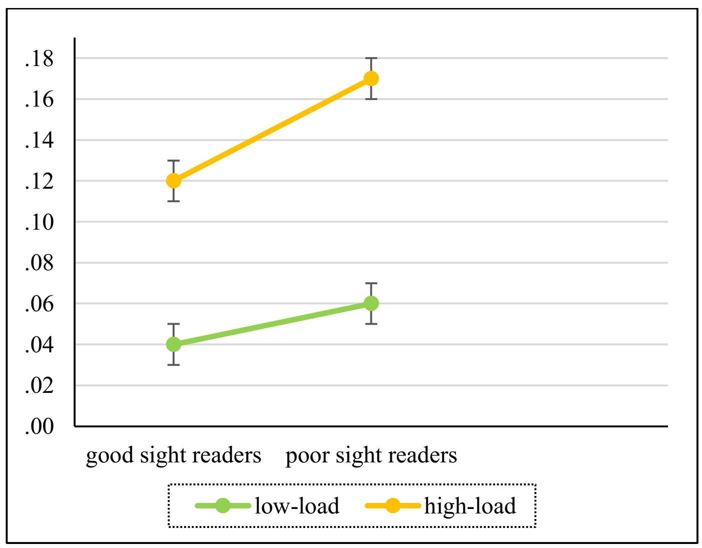


Figure 2. Proportion of same-list intrusions by group and memory load conditions.

3) Proportion of previous-lists intrusions.

The ANOVA showed a significant main effect for memory load (F (1, 74) = 17.822, p = .000, ηp^2 = .19), which means that the proportion of previous-lists intrusions was higher in the high load condition (4%) than in the low load condition (1%) for all the participants.

IV. DISCUSSION

Music performance at sight is a complex dual task that require the transcription of music printed in a score into sounds, in real time and without practice benefit. Taking into account the temporary nature of SR performance, we suggested that it could require the involvement of updating executive function sub-processes, that is, retrieval/transformation and substitution. Our first aim was to explore the relationship between the updating executive function sub-processes (retrieval/transformation and substitution) and SR rhythmic accuracy, a reliable measure of SR (Drake & Palmer, 2000; Elliot, 1982; Gromko, 2004; Hayward & Gromko, 2009; Kopiez & Lee, 2008; McPherson, 1994; Mishra, 2014; Waters et al., 1997; Wurtz et al., 2009) across development and expertise level. We hypothesized that, regardless of the effect of age and expertise, the processes of retrieval/transformation and substitution underlying updating executive function would be significantly related to SR rhythmic accuracy. Moreover, we also hypothesized that the differences between good and poor sight readers in terms of rhythmic performance could be due to updating sub-processes.

In relation to our first hypotheses, the results showed that the process of retrieval/transformation was significantly related to SR rhythmic accuracy in high load/low suppression condition, regardless of the age and level of expertise.

Thus, it seems that the ability to efficiently retrieve and transform information in WM is related to a better performance at sight, probably helping to select the adequate rhythmic proportion between notes from the first stages of music training.

Moreover, the results also showed that rhythmic accuracy was significantly related to inhibition in WM in high load/low suppression condition, but only in binary SR test. These pattern of correlations seems to indicate that the processes of SR and updating information in WM could be associated only in conditions of high memory and low suppression demands, that is, when WM was highly loaded but not overwhelmed.

With regard to our second hypothesis, globally, the results showed that good sight readers outperformed poor sight readers both in maintenance and inhibition in WM.

Specifically, and regarding the rhythmic accuracy in SR performance, our results showed that good sight readers had a greater capacity for maintaining information in WM than poor sight readers, regardless of their level of expertise and age, or the demands on memory load and suppression. These results could also support the implication of WM capacity in SR performance as Lehmann and Kopiez (2009) also suggested, at least in terms of rhythmic accuracy.

In relation to the substitution process, the updating task used in our study allowed us to differentiate between two processes: suppression of information in WM, that is, the ability to inhibit in WM information that is irrelevant to the goals of the task, and proactive interference, understood as the capacity to inhibit the information stored in LTM that is no longer relevant for the ongoing task. The results showed that good sight readers were less affected by the increase in memory load to inhibit irrelevant information from WM, whereas there were no significant differences between good and poor sight readers to inhibit information from LTM.

These results could be interpreted in terms of the limitations of the attentional system to simultaneously maintain and inhibit information in WM (Ecker et al., 2014). Thus, it is possible that good sight readers have a greater ability to inhibit irrelevant information in WM when memory load increases, allowing them to more efficiently maintain the relevant information.

In summary, we found that, in terms of rhythmic accuracy, good sight readers were significantly more efficient than poor sight readers both in retrieval/transformation and substitution processes underlying updating executive function, regardless of the effect of age and expertise level.

A limitation of this work is that SR performance must include not only rhythmic accuracy but also other measures such as pitch accuracy, continuity, articulation accuracy and expressiveness. It is possible that the relationship between updating sub-processes and these measures could reflect differences by age and expertise level.

ACKNOWLEDGMENT

We would like to thank Mª Teresa Lechuga for providing us with the word lists used in her experiment. We also thank Celia Viciana for their help in data collection, and the students who have collaborated with this research. This work was financially supported by MINECO EDU2011-22699 project.

REFERENCES

Barrouillet, P. (1996). Transitive inferences from set-inclusion relations and working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1408–1422. http://dx.doi.org/10.1037/0278-7393.22.6.1408

Barrouillet, P., & Lecas, J. (1999). Mental models in conditional reasoning and working memory. *Thinking & Reasoning*, 5, 289–302. http://dx.doi.org/10.1080/135467899393940

Belacchi, C., Carretti, B., & Cornoldi, C. (2010) The role of working memory and updating in Coloured Raven Matrices performance in typically developing children. *European Journal of Cognitive Psychology*,22:7,1010-1020, http://dx.doi.org/10.1080/09541440903184617

Bull, R., & Lee, K. (2014). Executive functioning and mathematics achievement. *Child Development Perspectives*, 8(1), 36-41. http://dx.doi.org/10.1111/cdep.12059

Carretti, B., Borella, E., Cornoldi, C., & De Beni, R. (2009). Role of working memory in explaining the performance of individuals with specific reading comprehension difficulties: A meta-analysis. *Learning and Individual Differences*, 19, 246-251.

Carriedo, N., Corral, A., Montoro, P., Herrero, L., & Rucián, M. (2016). Development of the updating executive function: from 7-year-olds to young adults. *Developmental Psychology*. Advance online publication. http://dx.doi.org/10.1037/dev0000091

Chen, T., & Li, D. (2007). The roles of working memory updating and processing speed in mediating age-related differences in fluid intelligence. *Aging, Neuropsychology, and Cognition*, 14(6), 631-646.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19(4), 450-466.

De Beni, R., & Palladino, P. (2004). Decline in working memory updating through ageing: Intrusion error analyses. *Memory*, 12(1), 75-89.

Drake, C., & Palmer, C. (2000). Skill acquisition in music performance: relations between planning and temporal control. *Cognition*, 74(1), 1-32.

Ecker, U. K., Lewandowsky, S., & Oberauer, K. (2014). Removal of information from working memory: A specific updating process. *Journal of Memory and Language*, 74, 77-90. http://dx.doi.org/10.1016/j.jml.2013.09.003.

Ecker, U. K., Lewandowsky, S., Oberauer, K., & Chee, A. E. (2010). The components of working memory updating: An experimental decomposition and individual differences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(1), 170-189.

Elliott, C. A. (1982). The relationships among instrumental sight-reading ability and seven selected predictor variables. *Journal of Research in Music Education*, 30(1), 5-14.

Gromko, J. E. (2004). Predictors of music sight-reading ability in high school wind players. *Journal of Research in Music Education*, 52(1), 6-15.

Hayward, C. M., & Gromko, J. E. (2009). Relationships among music sight-reading and technical proficiency, spatial visualization, and aural discrimination. *Journal of Research in Music Education*, 57(1), 26-36.

Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia*, 44, 2017–2036.http://dx.doi.org/10.1016/j.neuropsychologia.2006.01.010

Iglesias-Sarmiento, V., Carriedo-López, N., & Rodríguez-Rodríguez, J. L. (2014). Updating executive function and performance in reading comprehension and problem solving. *Anales de Psicología/Annals of Psychology*, 31(1), 298-309.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, 14(4), 389-433.

Kopiez, R. & Lee, J.I. (2006). Towards a dynamic model of skills involved in sight reading music. *Music Education Research*, 8 (1), 97–120.

Kopiez, R. & Lee, J. I. (2008). Towards a general model of skills involved in sight reading music. *Music Education Research*, 10 (1), 41–62.

Kopiez, R., C. Weihs, U. Ligges, and J.I. Lee. 2006. Classification of high and low achievers in a music sight reading task. *Psychology of Music* 34 (1), 5-26.

Lechuga, M. T., Moreno, V., Pelegrina, S., Gómez-Ariza, C. J., & Bajo, M. T. (2006). Age differences in memory control: Evidence from updating and retrieval-practice tasks. *Acta Psychologica*, 123(3), 279-298.

Lehmann, A. C., & Ericsson, K. A. (1996). Performance without preparation: Structure and acquisition of expert sight-reading and accompanying performance. *Psychomusicology: A journal of research in music cognition*, *15*(1-2), 1. http://dx.doi.org/10.1037/h0094082

Lehmann, A. C. & Kopiez, R. (2009). Sight-reading. In Hallam, S., Cross, I., & Thaut, M. (Eds.) *The Oxford handbook of music psychology*, pp. 344-351. Oxford: Oxford University Press.

Lehmann, A.C. & McArthur, V.H. (2002). Sight-reading. In Parncutt, R. & McPherson, G. (Eds) *Science and psychology of music performance*, pp. 135–150. Oxford University Press.

Lehto, J. E., Juujärvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology*, 21, 59 – 80. http://dx.doi.org/10.1348/ 026151003321164627

McPherson, G.E. 1994. Factors and abilities influencing sightreading skill in music. *Journal of Research in Music Education,* 42 (3), 217- 31.

McPherson, G. E., & Thompson, W. F. (1998). Assessing music performance: Issues and influences. *Research Studies in Music Education*, 10(1), 12-24.

Meinz, E. J., & Hambrick, D. Z. (2010). Deliberate practice is necessary but not sufficient to explain individual differences in piano sight-reading skill the role of working memory capacity. *Psychological Science*, 21, 914-919. http://dx.doi.org/10.1177/0956797610373933.

Mishra, J. (2014). Factors Related to Sight-Reading Accuracy A Meta-Analysis. *Journal of research in Music Education*, 61(4), 452-465.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, 41(1), 49-100.

Morris, N., & Jones, D. M. (1990). Memory updating in working memory: The role of the central executive. *British journal of psychology*, 81(2), 111-121.

Palladino, P., Cornoldi, C., De Beni, R., & Pazzaglia, F. (2001). Working memory and updating processes in reading comprehension. *Memory & Cognition*, 29(2), 344-354.

Passolunghi, M. C., & Pazzaglia, F. (2005). A comparison of updating processes in children good or poor in arithmetic word problem-solving. *Learning and Individual Differences*, 15(4), 257-269.

Sloboda, J. A. (1982). Music performance. In Deutch, D. (Ed.) *The psychology of music,* pp. 479-496. Academic press.

St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *The Quarterly Journal of Experimental Psychology*, 59, 745–759. http://dx.doi.org/10.1080/17470210500162854

Thompson, W. B. (1987). Music sight-reading skill in flute players. *The Journal of general psychology*, 114(4), 345-352.

Trinity College London (2003). *Sight reading pieces for Trumpet* (Grades 1-8). Sound and sight.

Trinity College London (2003). *Sight reading pieces for Violin* (Book 1: Initial-Grade 3; Book 2: Grades 4-8). Sound and sight.

Trinity College London (2004). *Sight reading pieces for Cello* (Initial-Grade 8). Sound and sight.

Trinity College London (2007). *Sight reading pieces for Bass Clef Brass* (Grades 1-8). Sound and sight.

Trinity College London (2007). *Sight reading pieces for Clarinet* (Book 1: Grades 1-4; Book 2: Grades 5-8). Sound and sight.

Trinity College London (2007). *Sight reading pieces for Flute* (Book 1: Grades 1-4; Book 2: Grades 5-8). Sound and sight.

Trinity College London (2007). *Sight reading pieces for Saxophone* (Book 1: Grades 1-4; Book 2: Grades 5-8). Sound and sight.

Trinity College London (2008). *Sight reading pieces for Oboe* (Grades 1-8). Sound and sight.

Trinity College London (2007). *Sight reading pieces for Viola* (Initial-Grade 8). Sound and sight.

Trinity College London (2009). *Sight reading pieces for Double Bass* (Initial-Grade 8). Sound and sight.

Trinity College London (2009). *Sight reading pieces for French Horn* (Grades 1-8). Sound and sight.

Waters, A.J., Townsend, E., & Underwood, G. (1998). Expertise in musical sight-reading: a study of pianists. *British Journal of Psychology*, 89, 123–149.

Waters, A. J., Underwood, G., & Findlay, J. M. (1997). Studying expertise in music reading: Use of a pattern-matching paradigm. *Perception & Psychophysics*, 59(4), 477-488.

Wolf, T. (1976). A cognitive model of musical sight-reading. *Journal of Psycholinguistic Research*, 5 (2), 143-171.

Wurtz, P., Mueri, R. M., & Wiesendanger, M. (2009). Sight-reading of violinists: Eye movements anticipate the musical flow. *Experimental Brain Research*, 194 (3), 445-450.

Xu, F., Han, Y., Sabbagh, M. A., Wang, T., Ren, X., & Li, C. (2013). Developmental differences in the structure of executive function in middle childhood and adolescence. PLoS ONE, 8(10),e77770 .http://dx.doi.org/10.1371/journal.pone.0077770

Auditory and visual beat and meter perception in children

Jessica E. Nave-Blodgett[1], Erin E. Hannon[1], and Joel S. Snyder[1]
[1]University of Nevada, Las Vegas. Las Vegas, Nevada, USA

Children experience music in their daily lives, and can tap along with the beat of the music, becoming more accurate with age. In Western cultures, music is often structured metrically, with beats theoretically heard as recurring patterns of stronger (downbeat) and weaker (upbeat) events. Do children perceive these so-called metrical hierarchies in music and match them with auditory or visual realizations of meter? Children aged 5-10 years old listened to excerpts of ballroom dance music paired with either auditory (beeping) or visual (ticking clock) metronomes. They rated how well the metronome matched the music they heard. For auditory metronomes, children "judged" a student drummer, and for visual metronomes, children "tested" musical clocks. Each child only experienced one metronome modality. There were four metronome conditions: synchronous with the music at the beat and measure level, synchronous at the level of the beat but not the measure, synchronous at the measure level but not the beat, or not at all synchronous with the music. For auditory metronomes, children at all ages successfully matched the beat level of the metronome with the music, and rated beat-synchronous metronomes as fitting the music better than beat-asynchronous metronomes. However, children did not use measure-level information in their judgments of fit, and they rated beat-synchronous metronomes as fitting equally well whether they also matched the measure level or not. In the visual modality, 5-6 year old children showed no evidence of beat or meter perception, and they did not differ in their ratings of any of the metronomes. Children aged 7-8 and 9-10 successfully used beat-level information in their ratings of fit, rating beat-synchronous metronomes as fitting better than beat-asynchronous visual metronomes. Beat perception appears to develop in the auditory modality earlier than the visual modality, and perception of multiple levels of metrical hierarchy may develop after age 10.

Beat-based entrainment during infant-directed singing supports social engagement

Miriam Lense[1,2,3] and Warren Jones[1,2]
[1]Marcus Autism Center, Children's Healthcare of Atlanta, Atlanta, GA, USA
[2]Department of Pediatrics, Emory University School of Medicine, Atlanta, GA, USA
[3]Program for Music, Mind, and Society, Vanderbilt University, Nashville, TN, USA

Social communication makes extensive use of predictable, rhythmic behaviors and individuals are acutely sensitive to the timing of these behaviors beginning in infancy. One example of such communication is infant-directed singing, a universal form of parent-child interaction that captures and maintains infants' attention and modulates their arousal levels. In the current study, we examined musical beat perception as a mechanism for the scaffolding of social attention and engagement. Specifically, we examined how entrainment to the beat directed attention to the eyes of others as eyes provide critical, meaningful social information (e.g., for directing attention or understanding emotions). Forty 24-month old toddlers watched videos of actresses engaging them with infant-directed singing while eye-tracking data was collected. We measured if changes in looking to the eyes of the actress were time-locked to the beat of the singing. Beats of the singing were determined through vowel onsets and offsets of the beat-aligned syllables. We constructed crosscorrelograms and compared these distributions with chance rates of eye-looking using permutation testing. Fixations to the eye regions of the actresses were time-locked to significantly increase in anticipation of the beat of the singing consistent with beat-based entrainment. These findings demonstrate how sensitivity to the beat can support meaningful engagement with a social partner during musical interactions in toddlers. Additionally, this research presents a novel method for examining entrainment to a musical beat in audiovisual stimuli that can be used with young children. Implications of this study for understanding the development of beat perception and sensitivity to social, rhythmic behaviors will be discussed.

Infants' home soundscapes: A portrait of American families

Eugenia Costa-Giomi[1]
[1]School of Music, The Ohio State University, Columbus, OH, USA

Infants learn complex aural communication systems rapidly and seemingly without effort simply through exposure to sounds. However, the way in which they learn to organize auditory information is dependent upon the characteristics of their soundscapes. For example, the type and quantity of infants' exposure to speech affect their language development and the musical experiences in which they are immersed affects their preferential attention and recognition of music. We studied infants' (n=8) language and music home environments by gathering precise and objective information about the characteristics, frequency, duration, and quality of infants' musical experiences during a weekend day through the use of an innovative technology that facilitates the collection and analysis of the sound environment. Results showed that live singing and purposeful music listening experiences did not occur frequently in the homes of middle class American families, that parents overestimated the number and length of such experiences in their reports, and that grandparents and infants' siblings often engaged in singing more regularly than did the infants' parents. Although infant-directed singing episodes occurred rarely, they were highly communicative, interactive, and intimate in nature. Overall, we found great diversity in the quantity and quality of the language and music experiences afforded by the infants. Some were exposed to recorded music played by the TV and digital toys for more than half of their waking hours, heard live singing for less than four minutes in an entire day, and were exposed to over 8000 words spoken by adults for a total of almost two hours. Others were immersed in more socially interactive language and musical experiences than others. Our finding that the quantity of infants' vocalizations was related to the characteristics of their soundcapes, highlights the importance of the home sound environment on infants' communicative skills.

Fetal facial expression in response to intravaginal music emission

Marisa López-Teijón[1], Alex García-Faura[1] and Alberto Prats-Gali[2]
[1]Institut Marquès, Barcelona, Spain
[2]Human Anatomy and Embryology Unit, Laboratory of Surgical Neuroanatomy, Facultat de Medicina, Universitat de Barcelona, Barcelona, Spain

This study compared fetal response to musical stimuli applied intravaginally (intravaginal music [IVM]) with application via emitters placed on the mother's abdomen (abdominal music [ABM]). Responses were quantified by recording facial movements identified on 3D/4D ultrasound. One hundred and six normal pregnancies between 14 and 39 weeks of gestation were randomized to 3D/4D ultrasound with: (a) ABM with standard headphones (flute monody at 98.6 dB); (b) IVM with a specially designed device emitting the same monody at 53.7 dB; or (c) intravaginal vibration (IVV; 125 Hz) at 68 dB with the same device. Facial movements were quantified at baseline, during stimulation, and for 5 minutes after stimulation was discontinued. In fetuses at a gestational age of >16 weeks, IVM-elicited mouthing (MT) and tongue expulsion (TE) in 86.7% and 46.6% of fetuses, respectively, with significant differences when compared with ABM and IVV. There were no changes from baseline in ABM and IVV. The frequency of TE with IVM increased significantly with gestational age. Fetuses at 16–39 weeks of gestation respond to intravaginally emitted music with repetitive MT and TE movements not observed with ABM or IVV. Our findings suggest that neural pathways participating in the auditory–motor system are developed as early as gestational week 16. These findings might contribute to diagnostic methods for prenatal hearing screening, and research into fetal neurological stimulation. Presented by Professor Jordi A. Jauset-Berrocal, Ph D (University Ramon Llull, Barcelona, Spain), consultant in this research. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4616906/pdf/10.1177_1742271X15609367.pdf.

Fourteen- but not twelve-month-olds use interpersonal synchrony as a cue for third-party relationships

Laura Cirelli[1], Stephanie Wan[1], Laurel Trainor[1,2,3]
[1]Department of Psychology, Neuroscience & Behaviour, McMaster University, Canada
[2]McMaster Institute for Music and the Mind, McMaster University, Canada
[3]Rotman Research Institute, Baycrest Hospital, Toronto, Ontario, Canada

Adults expect synchronously- versus asynchronously-moving individuals to be higher in rapport. This third-party social expectation is especially interesting since experiencing interpersonal synchrony encourages prosociality. The influence of interpersonal synchrony on social behaviour develops early. More specifically, 14-month-old infants display more helpfulness towards a synchronously- compared to asynchronously-moving experimenter. Additionally, 12-month-olds are more likely to select a synchronously-over asynchronously-moving stuffed toy. The present study investigated whether interpersonal synchrony observed between third-party individuals cues expectations for affiliation in 12- (n=16) and 14-month old (n=32) infants. Infants sat on their parent's lap and watched videos of pairs of women moving either synchronously or asynchronously to music for 2 minutes. Afterwards, infants watched test trial videos of these women interacting either positively (waving) or negatively (turning away). Looking times were measured to assess infant expectations about how these women should interact. Because infants look longer at surprising events, we predicted that they would look longer to negative interactions between synchronous partners, and to positive interactions between asynchronous partners. We found no effect of movement condition or test trial type on test trial looking times for 12-month-olds. However, we found the predicted interaction in 14-month-olds. They were surprised to see asynchronous partners interacting positively, suggesting that they did not expect these women to be affiliates. However, there was no difference in looking times to either test trial in the synchrony conditions. These results show that 14- but not 12-month old infants have different expectations about third-party social relationships between those who move in- or out-of-synchrony with one another.

Investigating the importance of self-theories of intelligence and musicality for students' academic and musical achievement

Daniel Müllensiefen[1], Peter Harrison[1], Francesco Caprini[1] and Amy Fancourt[2]
[1]Department of Psychology, Goldsmiths, University of London, London UK
[2]Brain Can Do, Queen Anne's School, Reading, UK

Musical abilities and active engagement with music have been shown to be positively associated with many cognitive abilities as well as social skills and academic performance in secondary school students. While there is evidence from intervention studies that musical training can be a cause of these positive relationships, recent findings in the literature have suggested that other factors (e.g. genetics, family background, personality traits) might also contribute. In addition, there is mounting evidence that self-concepts and beliefs can affect academic performance independently of intellectual ability. However, it is currently not known whether student's beliefs about the nature of musical talent also influence the development of musical abilities in a similar fashion. Therefore, this study introduces a short self-report measure termed "Musical Self-Theories and Goals," closely modeled on validated measures for self-theories in academic scenarios. Using this measure the study investigates whether musical self-theories are related to students' musical development as indexed by their concurrent musical activities and their performance on a battery of listening tests. We use data from a cross-sectional sample of 313 secondary school students to construct a network model describing the relationships between self-theories and academic as well as musical outcome measures, while also assessing potential effects of intelligence and the Big Five personality dimensions. Results from the network model indicate that self-theories of intelligence and musicality are closely related. In addition, both kinds of self-theories are connected to the students' academic achievement through the personality dimension conscientiousness and academic effort. Finally, applying Pearl's do-calculus method to the network model we estimate that the size of the assumed causal effects between musical self-theories and academic achievement lie between 0.07 and 0.15 standard deviations.

Musical Sophistication and Musical Self-Concept Influence the Different Trajectories of Musical Development in Secondary School Students

Daniel Fiedler,[*1] Daniel Müllensiefen,[#2]

[*]*Institute of Music, University of Education Freiburg, Germany*
[#]*Goldsmiths., University of London, United Kingdom*
[1]`daniel.fiedler@ph-freiburg.de`, [2]`d.mullensiefen@gold.ac.uk`

ABSTRACT

Trajectories of musical development can be very different across adolescence and the causes and mechanisms leading to these differences are often subjects of music educational research. To measure the aspects of musical development of German students, the psychometric constructs *musical self-concept* and *musical sophistication* can be used. The first aim of this repeated-measurement study was to identify different developmental trajectories within the sample of students as well as to investigate associations of two typical trajectories with socio-demographic variables. The second aim was to identify how *musical self-concept* and *musical sophistication* contribute to the development of the target variable *interest in music as a school subject* for the both identified developmental trajectories. Data of 334 students (f = 188, m = 146) from three Grammar Schools (n = 223), three Middle Schools (n = 90) and one Junior High School (n = 21) are presented. The data comprised the self-assessed psychometric constructs as well as music-specific and demographic background variables and the target variable of this study at three time points. The data were analyzed using SEMs, a sequence pattern analysis and multilevel models. The sequence pattern analyses identified three respectively four developmental trajectories for the constructs *musical self-concept* and *musical sophistication*. In addition, significant correlations between two typical developmental trajectories with the variables sex (φ = .319), musical status (φ = .416) and type of school (φc = .222) are found. Moreover, a multilevel analysis shows that the overall *interest in music as a school subject* is decreasing over time. But for the identified higher trajectory group the target variable *interest in music* is increasing over time. Hence, this study makes an important contribution to the understanding of the mechanisms of musical development during adolescence.

I. INTRODUCTION

Musical expertise, skills and behaviours can develop very differently during adolescence, and the reasons and mechanisms behind these differences are often objects of research in music psychology and music education. It is generally assumed that the development of musical skills is integrated into the individual's overall development, where culture and education system (Gembris, 2013) as well as interest and motivation (e.g. *interest in music as school subject*) play important roles (Heß, 2011a). In the past, musical skills, expertise, and competences either were measured with tests that quantify students' musical competences and achievements (Colwell, 1969; 1970; 1979; Gordon, 1971; Bähr, 2001), or were measured with aptitude and musicality tests that assess the student's potential for future musical achievement (Seashore, 1919; Bentley, 1968; Gordon, 1989). Furthermore, several German and international longitudinal surveys have investigated the (positive) influence of musical training on non-musical abilities and behaviours, such as intelligence, social behaviours and cognitive effects, (e.g. Weber, Spychiger & Party, 1993; Bastian, 2000; Ho, Cheung & Chan, 2003; Schellenberg, 2004). However, in music psychology (and sometimes in music education) studies, the measure of musical skills, expertise, and competences as well as experiences and identities is often simply the participant's amount of instrumental musical training or extracurricular music education. Using such simplified measures neglects the complex and multi-faceted nature of musical expertise, skill, and related behaviour (Gembris, 2013; Hallam, 2010; 2006; Hallam & Prince, 2003).

II. AIM OF THIS STUDY

The first aim of this repeated-measurement study was to identify different developmental trajectories within the sample of students by using the Goldsmiths Musical Sophistication Index (Gold-MSI; Fiedler & Müllensiefen, 2015; Schaal et al., 2014; Müllensiefen et al., 2014) as well as the Musical Self-Concept Inquiry (MUSCI; Spychiger, 2010; 2012; Spychiger & Hechler, 2014). Both questionnaires were developed to measure different facets of *musical sophistication* (Gold-MSI) as well as *musical self-concept* (MUSCI) of students at secondary education schools. The second aim of this study was to investigate associations of typical trajectories with socio-demographic variables such as *sex*, *musical status*, and *type of school*. The third aim was to identify how *musical self-concept* and *musical sophistication* contribute to the development of the target variable of this study *interest in music as a school subject* for two typical identified developmental trajectories.

III. GOLD-MSI AND MUSCI QUESTIONNAIRE

A. Gold-MSI Questionnaire to Measure *Musical Sophistication*

In line with Hallam (Hallam & Prince, 2003; Hallam, 2010) and Ollen (2006), the novel term *musical sophistication* conceptually includes different facets of musical expertise and behaviours. Furthermore, it "has been used infrequently in earlier research and is therefore less loaded with biases and preconceptions than more commonly used terms such as musicality, musical talent, ability, aptitude, or musical potential" (Müllensiefen et al., 2014, p. 2). This conceptualisation of *musical sophistication* assumes that musical abilities, skills, and behaviours are developed through active engagement with music in its many different forms as well as that individual differences in the sophistication of observable musical behaviours are related to differences in cognitive categorisation and processing of music (Müllensiefen et al., 2014; Schaal et al., 2014). High levels of *musical sophistication* "are generally characterized by a) higher frequencies of exerting musical skills or behaviours, b) greater ease, accuracy or effect of musical behaviours when executed, and c) a greater and more varied repertoire of musical behaviour patterns" (Müllensiefen et al., 2014, p. 2). So the empirically derived differentiation of *musical sophistication* contains the five subscales *Active Engagement (F1)*, *Perceptual Abilities (F2)*, *Musical Training (F3)*, *Singing Abilities (F4)* and *Emotions (F5)* as well as the general factor *Musical Sophistication (F6)*, and therefore describes musical behaviours "ranging from performance on an instrument and listening expertise, to the ability to employ music in functional settings or to communicate about music" (Müllensiefen et al., 2014, p. 1).

B. MUSCI Questionnaire to Assess *Musical Self-Concept*

Following the seminal work by Shavelson, et al. (1976) the self-concept presents hierarchically ordered layers and a number of parts. "Layers as well as their parts are called domains of the self-concept. While the domains within a layer are similar with regard to their formal value, they distinctively differ from one another with regard to their content" (Spychiger, in press, p. 269). This means, that "the first layer distinguishes between the academic and the non-academic self-concept, and the second between their sub-domains" (Spychiger, in press, p. 269). With this in mind, musical self-concept is one part of a person's general self-concept (Spychiger, in press; 2007, 2010, 2012). In addition, the term *musical self-concept* is hypothesized to be the psychological structure that turns personal musical experiences into musical identity. Connecting musical identity to *musical self-concept* sheds light on basic activities of the mind, awareness and consciousness, and on the interactive concept of recognition (Spychiger, in press). So "musical self-concept summarizes a person's answers to his or her inquiries into 'who-I-am' and 'what-I-can-do' questions with regards to music" (Spychiger, in press, p. 268). Above this, *musical self-concept* includes different facets of ideas, perceptions, and assessments – the cognitions – a person has regarding its own musical activities (Bernecker, Haag & Pfeiffer, 2006, p. 53).

To measure *musical self-concept* a multilevel process was used to empirically operationalize the construct of *musical self-concept* (Spychiger, 2010) by using different German-speaking samples of adults. The result of this operationalization process is the MUSCI questionnaire, with its empirically derived multidimensional factor structure, assessing many different elements of musical experiences and musical identity (Spychiger, 2007, 2010, 2012; Spychiger & Hechler, 2014; Spychiger, in press). The statistical results of the three waves of investigation are 63 items developed using reliability and factor analyses. The final questionnaire was named MUSCI, the Musical Self-Concept Inquiry (Spychiger et al., 2009; 2010), and comprises 12 scales, of which 4 scales are to be completed by musicians (person who currently play a musical instrument) only. The MUSCI questionnaire, which is used in this study, to assess *musical self-concept* of musically non-active and active students, comprises the eight factors *Mood Management (S1)*, *Community (S2)*, *Technique & Information (S3)*, *Musical Ability (S4)*, *Movement & Dance (S5)*, *Spirituality (S6)*, *Ideal Music Self (S7)*, and *Adaptive Music Self (S8)* (with altogether 43 items; for more information see Spychiger, in press).

IV. METHOD

A. Sample

The sample consisted of 334 students (female = 180, male = 146) from three Grammar (n = 223) and three Middle (n = 90) Schools as well as one Junior High School (n = 21) across different regions in the south-west of Germany (Baden-Württemberg). The average age at the beginning of this study was 12.40 years (SD = 1.50) with an age range of 9 to 17 years. Concerning the age and types of school as well as teaching groups the sample shows no representativeness.

B. Data Collection and Measurement Instruments

The complete questionnaire was distributed on paper and assessed *musical self-concept* with the MUSCI questionnaire (4-point Likert-scale), *musical sophistication* with the Gold-MSI (7-point Likert-scale), *self-closeness to 'music'* (Kessels & Hannover, 2004; Heß, 2011b), *interest in the school subject 'music'* (Rakoczy, Klieme & Pauli, 2008), and *self-attribution concerning marks in 'music'* (Rakoczy, Buff & Lipowsky, 2005, S. 164).

C. Data Analyses

The data were analyzed using confirmatory factor analysis (CFA) to verify the factor structures of the MUSCI as well as Gold-MSI questionnaire and to assess the Factor Reliability (FR) and Average Variance Extracted (AVE). Reliability measurements were employed to determine the validity (Cronbach's Alpha), correlational analyses to assess the criterion-related validity, and structural equation modeling (SEM) to reveal the development of the target variable *interest in music as a school subject*. Moreover, the data were analyzed using a sequence pattern analysis to identify different developmental trajectories within the students' sample and multilevel models.

V. RESULTS

As Fiedler and Müllensiefen (in press) already showed, test-retest-reliability and Cronbach's Alpha indicate acceptable to good subscale reliabilities for the psychometric constructs *musical self-concept* and *musical sophistication*. In addition, several structural equation models indicate the relationship between the sub-facets of *musical self-concept* and *musical sophistication* with the target variable *interest in music as a school subject* as well as relationships with other music-specific and demographic background variables – e.g. age, *self-closeness to music as a school subject*, *self-assessed marks in the school subject music* – (Fiedler & Müllensiefen, in press). However, there is no study, which shows that there are different developmental trajectory groups of the psychometric constructs *musical self-concept* and *musical sophistication*, which differ from the absolute level as well as them continually changing over time.

With the help of the R-Package TraMineR (Gabadinho, Ritschard, Studer & Müller, 2011; Stegmann, Werner & Müller, 2013) a sequence pattern analysis was conducted to identify the different musical trajectory groups within the sample of students. Thereby, the sequence pattern analysis is a method to identify different person clusters in longitudinal surveys

(Aisenbrey & Fasang, 2010). As a result, the sequence pattern analysis identified three, respectively four different trajectories for the sub-facets of the psychometric constructs *musical self-concept* and the general factor of *musical sophistication.* In order to analyze relationships, the correlations between the three, respectively four identified trajectory groups with the assessed socio-demographic background variables were examined. As a result, table 1 and 2[1] show the correlations between the identified developmental trajectories of the sub-facets of *musical self-concept* and the general factor of *musical sophistication* with students' *sex*, *musical status*, and *type of school*, with significant phi-correlation-coefficients between φ = .247 and φ = .459 – for the identified four developmental trajectories of the general factor of *musical sophistication* – as well as significant phi-correlation-coefficients between φ = .297 and φ = .358 – for the three developmental trajectories of the sub-facets of *musical self-concept*. Thereby, one exception can be found, because there is no significant relationship between the developmental trajectories of the sub-facets of *musical self-concept* with *type of school*. Moreover, the analysis shows that there are two typical developmental trajectory groups, which can be distinguished. In addition, significant correlations between the two typical identified developmental trajectory groups with the variables sex (φ = .319), musical status (φ = .416), and type of school (φ_c = .222) are found. However, the examined significant correlations show that female as well as musically active[2] students and especially students at Grammar Schools are associated with higher developmental trajectories.

In addition, to identify how *musical self-concept* and *musical sophistication* contribute the development of the target variable of this study interest in music as a school subject the two typical identified developmental trajectory groups were used to conduct a multilevel analysis. As a result, the multilevel analysis shows that the overall *interest in music as a school subject* (target variable) is decreasing over time ($p \leq .001$). But in addition, the analysis shows that there is also an interaction effect. This means that for the students in the higher trajectory group, the target variable *interest in music as a school subject* is increasing over time ($p \leq .001$), compared to their peers in the lower trajectory group.

VI. DISCUSSION

The first aim of this repeated-measurement study was to identify different developmental trajectories within the sample of students by using the Goldsmiths Musical Sophistication Index (Gold-MSI; Fiedler & Müllensiefen, 2015; Schaal et al., 2014; Müllensiefen et al., 2014) as well as the Musical Self-Concept Inquiry (MUSCI; Spychiger, 2010; 2012; Spychiger & Hechler, 2014). As a result, three, respectively four different developmental trajectories for the sub-facets of *musical self-concept* (MUSCI) and the general factor of *musical sophistication* (Gold-MSI) were identified. Moreover, correlations between the identified developmental trajectory groups with the assessed socio-demographic background variables such as students' *sex*, *musical status*, and *type of school* were analyzed (see table 1 and 2). In addition, two typical developmental trajectory groups could be distinguished. As a result of the correlational analyses, the significant phi-correlation-coefficients show that *sex* (female students) and *musical status* (musically active students) as well as *type of school* (Grammar Schools) are associated with higher developmental trajectories of the psychometric constructs as well as the higher developmental trajectory group. This means, that especially musically active female students at Grammar Schools are in higher developmental trajectory groups and show a higher development of *musical self-concept* as well as *musical sophistication*, compared to their male peers at lower musical developmental trajectory groups. Thus, the identified relationships suggest that school music lessons can have a non-inclusive character – particular for male students and students who don't play a musical instrument– (Heß, 2011a; 2013), which particular can influence the development of the target variable *interest in music as a school*. Therefore, it is not surprising that the multilevel analysis shows that the overall *interest in music as a school subject* is decreasing over the three time points in the school year 2014/2015. This result is in line with Daniels' (2008, p. 348) outcomes. Daniels (2008) describes that the interest in a school subject is decreasing at class seven (p. 348). The reasons for this are among other things a lack between the (music) lessons in school and the psychological basic human needs for autonomy, social integration, and competency (Daniels, 2008, p. 348; Deci & Ryan, 1985; 2002). In summary, *musical self-concept* as well as *musical sophistication* and the identified developmental trajectory groups of the assessed psychometric constructs contribute the development of the target variable *interest in music as a school subject* over time. This means that for the higher developmental trajectory group the target variable *interest in music as a school subject* is increasing over time, compared to their peers in the lower developmental trajectory group. Hence, this study makes an important contribution to the understanding of the mechanisms of musical development during adolescence. Moreover, the results of this study can serve as a basis for teaching segmentation by ability and interest in music school teaching.

[1] Note that table 1 and 2 can be found at the end of this paper.
[2] Musically active students are persons who currently play a musical instrument.

REFERENCES

Aisenbrey, S. & Fasang, A. E. (2010). New Life for Old Ideas: The "Second Wave" of Sequence Analysis. Bringing the "Course" Back Into the Life Course. *Sociological Methods & Research, 38(3)*, 420-462.

Bähr, J. (2001). *Zur Entwicklung musikalischer Fähigkeiten von Zehn- bis Zwölfjährigen: Evaluation eines Modellversuchs zur Kooperation von Schule und Musikschule*. Göttingen: Cuvillier.

Bastian, H. G. (2000). *Musik(erziehung) und ihre Wirkung: Eine Langzeitstudie an Berliner Grundschulen*. Mainz: Schott.

Bentley, A. (1968). *Musikalische Begabung bei Kindern und ihre Meßbarkeit*. Frankfurt: Diesterweg.

Bernecker, C., Haag, L. & Pfeiffer, W. (2006). Musikalisches Selbstkonzept. Eine empirische Untersuchung. *Diskussion Musikpädagogik, 29*, 53-56.

Colwell, R. (1969). *Music achievement tests 1 and 2*. Chicago: Follett Educational Corp.

Colwell, R. (1970). *Music achievement tests 3 and 4*. Chicago: Follett Educational Corp.

Colwell, R. (1979). *Silver Burdett Music Competency Tests*. Morristown: Silver Burdett.

Daniels, Z. (2008). *Entwicklung schulischer Interessen im Jugendalter*. Münster: Waxmann.
Fiedler, D. & Müllensiefen, D. (2015). Validierung des Gold-MSI-Fragebogens zur Messung Musikalischer Erfahrenheit von Schülerinnen und Schülern der Sekundarstufen an allgemeinbildenden Schulen. In A. Niessen & J. Knigge (Eds.), *Theoretische Rahmung und Theoriebildung in der musikpädagogischen Forschung* (pp. 199-219). Musikpädagogische Forschung: Bd. 36. Münster: Waxmann.
Fiedler, D. & Müllensiefen, D. (in press). Structure and development of musical self-concept, musical sophistication and interest in the school subject 'music'. An empirical long-term study of German students (9 to 17 years) at secondary modern, inter-denominational and middle schools as well as grammar schools (in German).
Gabadinho, A., Ritschard, G., Struder, M. & Müller, N. S. (2011). *Minningsequence data in R with the TraMineR package. A user's guide.* University of Geneva. Retrieved April, 28, 2016, from http://mephisto.unige.ch/pub/TraMineR/doc/TraMineR-Users-Guide.pdf.
Gembris, H. (2013). *Grundlagen musikalischer Begabung und Entwicklung* (4th ed.). Augsburg: Wißner.
Gordon, E. E. (1971). *Iowa Test of Music Literacy*. Iowa: The Bureau of Educational Research and Services, University of Iowa.
Gordon, E. E. (1989). *Advanced Measures of Music Audiation*. Chicago: GIA.
Hallam, S. (2006). 'Musicality'. In G McPherson (Ed.), *The child as musician: A handbook of musical development* (pp. 93-110). Oxford: Oxford University Press.
Hallam, S. (2010). 21st century conceptions of musical ability. *Psychology of Music, 38(3)*, 308-330.
Hallam, S., & Prince, V. (2003). Conceptions of Musical Ability. *Research Studies in Music Education, 20(1)*, 2-22.
Heß, F. (2011a). Musikunterricht zwischen Sach- und Fachinteresse. Ergebnisse aus der Pilotstudie Musikunterricht aus Schülersicht. *Beiträge empirischer Musikpädagogik, 2(1)*, 1-26. Retrieved April, 28, 2016, from http://www.b-em.info/index.php?journal=ojs&page=editor&op=proofGalley&path%5B%5D=44&path%5B%5D=102.
Heß, F. (2011b). Skalenhandbuch zur Studie „Musikunterricht aus Schülersicht" (MASS 2011). Dokumentation der Erhebungsinstrumente. Retrieved April, 28, 2016, from https://www.uni-kassel.de/fb01/fileadmin/datas/fb01/Institut_fuer_Musik/Dateien/Skalenhandbuch_Mass_2011.pdf.
Heß, F. (2013). Musikunterricht - ein Mädchenfach? Das Image des Faches Musik aus Sicht von Jugendlichen. *Musik & Unterricht, 110*, 56-60.
Ho, Y.-C., Cheung, M.-C., & Chan, A. S. (2003). Music Training Improves Verbal but Not Visual Memory: Cross-Sectional and Longitudinal Explorations in Children. *Neuropsychology, 17(3)*, 439-450.
Kessels, U., & Hannover, B. (2004). Empfundene „Selbstnähe" als Mediator zwischen Fähigkeitsselbstkonzept und Leistungskurswahlintentionen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 36(3)*, 130-138.
Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Publication. *PLoS ONE, 9(2)*: e89642. doi:10.1371/journal.pone.0089642.
Ollen, J. E. (2006). *A criterion-related validity test of selected indicators of musical sophistication using expert ratings*. Dissertation. Ohio State University: Ohio.
Rakoczy, K., Buff, A., & Lipowsky, F. (2005). *Dokumentation der Erhebungs- und Auswertungs-instrumente zur schweizerisch-deutschen Videostudie „Unterrichtsqualität, Lernverhalten und mathematisches Verständnis". Teil 1 Befragungsinstrumente*. Bd. 13. Retrieved April, 28, 2016 from http://www.pedocs.de/volltexte/2010/3106/pdf/MatBild_Bd13_D_A.pdf.
Rakoczy, K., Klieme, E., & Pauli, C. (2008). Die Bedeutung der wahrgenommenen Unterstützung motivationsrelevanter Bedürfnisse und des Alltagsbezugs im Mathematikunterricht für die selbstbestimmte Motivation. *Zeitschrift für Pädagogische Psychologie, 22(1)*, 25-35.
Schaal, N. K., Bauer, A.-K. R., & Müllensiefen, D. (2014). Der Gold-MSI: Replikation und Validierung eines Fragebogeninstrumentes zur Messung Musika-lischer Erfahrenheit anhand einer deutschen Stichprobe. *Musicae Scientiae, 18(4)*, 423-447. doi: 10.1177/1029864914541851.
Schellenberg, E. G. (2004). Music Lessons Enhance IQ. *Psychological Science, 15(8)*, 511-514.
Seashore, C. E. (1919). *Seashore measurement of musical talent*. New York, rev. Ed. 1939, 1956, 1960. German Edition, H. Fischer & C. Butch (Eds.). Bern: Huber.
Shavelson, R. J., Hubner, J. J. & Stanton, G. C. (1976). Self-Concept: Validation of Construct Interpretations. *Review of Educational Research, 46(3)*, 407-441.
Spychiger, M. (2007). „Nein, ich bin ja unbegabt und liebe Musik". Ausführungen zu einer mehrdimensionalen Anlage des musikalischen Selbstkonzepts. *Diskussion Musikpädagogik, 33*, 9-20.
Spychiger, M. (2010). *Das musikalische Selbstkonzept. Konzeption des Konstrukts als mehrdimensionale Domäne und Entwicklung eines Messverfahrens. Schlussbericht an den Schweizerischen Nationalfonds zur Förderung der Wissenschaften*. Frankfurt am Main: Hochschule für Musik und Darstellende Kunst, Fachbereich 2 (unpublished).
Spychiger, M. (2012). Das musikalische Selbstkonzept. Drittmittelfinanzierte musikpsycho-logische Grundlagenforschung an der HfMDK Frankfurt am Main. *Frankfurt in Takt. Magazin der Hochschule für Musik und Darstellende Kunst Frankfurt am Main, 12(2)*, 43-47.
Spychiger, Maria (in press). Musical self-concept as a mediating psychological structure. From musical experience to musical identity. in D.J. Hargreaves, R. MacDonald & D. Miell (Eds.): *The Oxford Handbook on Musical Identity* (pp. 267-287). Oxford: Oxford UP.
Spychiger, M., Gruber, L. & Olbertz, F. (2009). Musical Self-Concept. Presentation of a Multi-Dimensional Model and Its Empirical Analyses. In J. Louhivuori, T. Eerola, S. Saarikallio, T. Himberg & P.-S. Eerola (Eds.), *Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music (ESCOM 2009)* (pp. 503-506). Jyväskylä, Finland.
Spychiger, M. & Hechler, J. (2014). Musikalität, Intelligenz und Persönlichkeit. Alte und neue Integrationsversuche. In W. Gruhn & A. Seither-Preisler (Eds.), *Der musikalische Mensch. Evolution, Biologie und Pädagogik musikalischer Begabung* (S. 23-68). Hildesheim, Zürich, New York: Olms.
Weber, E. W., Spychiger, M., & Patry, J.-L. (1993). *Musik macht Schule. Biografie und Ergebnisse eines Schulversuchs mit erweitertem Musikunterricht*. Essen: Blaue Eule.

Table 1: Overview of the musical developmental trajectory groups of the sub-facets of musical self-concept as well as the phi-correlation-coefficients to describe the relationships between the musical developmental trajectory groups with the socio-demographic background variables sex and musical status

musical developmental trajectory group of the sub-facets of musical self-concept	***sex***		***musical status***		***type of school***		
	female	*male*	*non-active*	*active*	*Grammar School*	*Middle School*	*Junior High School*
group 1 - "middle" (n = 156)	82	74	78	78	103	40	13
group 2 - "middle to high" (n = 136)	99	37	38	98	96	36	4
group 3 - "low" (n = 42)	7	35	30	12	24	14	4
total	**188**	**146**	**146**	**188**	**223**	**90**	**21**
phi-coefficients (φ)	.358 ($p \leq .001$)		.297 ($p \leq .001$)		-		

Table 2: Overview of the musical developmental trajectory groups of the general factor of musical sophistication as well as the phi-correlation-coefficients to describe the relationships between the musical developmental trajectory groups with the socio-demographic background variables sex, musical status, and type of school

musical developmental trajectory group of the general factor of musical sophistication	***sex***		***musical status***		***type of school***		
	female	*male*	*non-active*	*active*	*Grammar School*	*Middle School*	*Junior High School*
group 1 - "low to middle" (n = 90)	40	50	48	42	56	27	7
group 2 - "low" (n = 76)	27	49	59	17	39	27	10
group 3 - " middle to high" (n = 70)	50	20	11	59	53	14	3
group 4 - "high" (n = 98)	71	27	28	70	75	22	1
total	**188**	**146**	**146**	**188**	**223**	**90**	**21**
phi-coefficients (φ)	.325 ($p \leq .001$)		.459 ($p \leq .001$)		.247 ($p \leq .001$)		

Advancing Interdisciplinary Research in Singing (AIRS): Early Development

Mayumi Adachi[1], Stefanie Stadler Elmer[2]
[1] Department of Psychology, Hokkaido University, Sapporo, Hokkaido, Japan, m.adachi@let.hokudai.ac.jp
[2] Department of Psychology, University of Zurich, Zurich Switzerland

Singing begins spontaneously early in life, as the case for speech. Although there has been extensive research on children's acquisition of speech, relatively little attention has been devoted to the acquisition of singing, especially in the early years. Even within music psychology, the primary focus on skill acquisition has been on learning to play a musical instrument rather than learning to perform on one's inborn vocal instrument. Singing entails multiple skills that include vocal play and gesture as well as learning and reproduction of melodies (pitch and rhythmic patterns) and lyrics. The present symposium, under the auspices of the AIRS (Advancing Interdisciplinary Research in Singing) Project, includes presentations by Anna Rita Addessi (Italy) on infants, Sandra Trehub (Canada) on toddlers, and Beatriz Ilari (U.S.A) on children of primary school age, with Mayumi Adachi (Japan) and Stefanie Stadler Elmer (Switzerland) as discussants. Anna Rita Addessi examines vocal interactions between caregivers and infants that occur during routine diaper-changing. Sandra Trehub's analysis of home recordings of toddlers' singing (from YouTube and parents) reveals that toddlers are capable of reproducing the contours, rhythms, and pitch range of conventional songs. Beatriz Ilari examines the long-term impact of musical (e.g., Il Systema) and nonmusical (athletic) interventions on various singing behaviors. Although children's improvisation abilities increased substantially over the three years of the study, the musical interventions proved to be no more effective than the nonmusical interventions. The presentations and discussion provide novel findings across a wide age range, novel methods, and novel perspectives from different disciplines (music education, psychology) settings (home, school, community), and countries.

Vocal development in infancy: The role of reflexivity during interactions with parents, grandmother, and educator during routine care

Anna Rita Addessi
[1]Department of Educational Sciences, University of Bologna, Bologna, Italy

This study takes a constructionist perspective concerning the interaction of an infant with various adults. It focuses in particular on an echo mechanism of repetition and variation, which structures the child-adult interaction according to different time scales (micro-events to daily rhythms), languages, modes of expression, and communication. A previous case-study with one infant 8 months of age was videorecorded at home during the daily diaper change on successive days, with mother and father. The observed relation between the infant and adult behavior highlighted the importance of imitation and variation in these dyadic interchanges and several differences between mother and father. The present study reports on a larger study using this same paradigm. Eleven infant 8 months of age were videorecorded during the daily diaper change on 7 successive days, at home and in the day-care nursery, providing 7 recordings each of the infant with the mother, father, grandmother, and educator. Audiovideo data was scored via a grid with respect to duration and frequency of vocal productions, imitation, variation, and turn taking. The results showed that the child's vocal activity is inversely proportional to the adult's vocal activity (duration and frequency) and increases when the adult imitates the child and respects the turn-taking. Several differences between mother, father, grandmother and educator were observed concerning the intentionality, the timing of the interaction, and the musical quality of the vocalisations. Data analysis now in progress will be presented and discussed. The results observed to date highlight the importance of the reflexive interaction between child and adult and can have important implications in the field of music education. Because the diaper activity takes place daily for every infant, these results have great generality and highlight the influences on infant vocal learning in everyday activities.

Toddlers' Singing

Sandra E. Trehub[1]
[1]Department of Psychology, University of Toronto, Toronto, Ontario, Canada

According to most descriptions, toddlers focus on song lyrics rather than melodies when they sing, and the resulting songs have a relatively narrow pitch range. However, case studies have revealed more proficient singing in toddlers, raising the possibility that those toddlers were precocious rather than representative. Another possibility is that toddlers' spontaneous singing at home may reveal more about their capabilities than their singing of experimenter-selected songs in less familiar environments. In one study involving home recordings from toddlers 18 to 36 months of age, we found a wide range of singing proficiency that was unrelated to speech production skills. At 20 months, a number of toddlers were merely producing the words of songs in a monotone voice. By 26 months, however, most toddlers reproduced the pitch contours, rhythms, and words of songs, with large individual differences. In another study (Helga Gudmundsdottir, collaborator), we sought to ascertain whether toddlers' sung melodies were recognizable when sung with foreign words. We gathered home recordings of toddlers 16 months to 3 years (from YouTube and parents). We selected several samples of two familiar songs (*Twinkle, Twinkle; Happy Birthday*) and two unfamiliar children's songs, all sung in foreign languages. In online testing, English-speaking adults were required to identify the tunes. They readily identified the familiar tunes (over 90% correct) but not the unfamiliar tunes. These findings disconfirm the claim that toddlers' songs are recognizable from the lyrics. Pitch analyses revealed that toddlers' pitch range approximated that of the target songs, disconfirming the claims of reduced pitch range. Finally, in an ongoing study, we are attempting to ascertain whether toddlers' proficiency in imitating visual gestures is related to their proficiency in singing, which relies, in part, on vocal imitation.

From imitation to creation: Children's developing improvisatory skills

Beatriz Ilari[1], Cara Fesjian[2], Bronte Ficek[2] and Assal Habibi[2]
[1]Department of Music Teaching & Learning, University of Southern California, Los Angeles, CA, USA
[2]Brain & Creativity Institute, University of Southern California, Los Angeles, CA, USA

Improvisation is one important means for developing creative thinking in music. In spite of much discourse in support of improvisation in classrooms, little is currently known about the development of children's improvisatory skills. The aim of this mixed methods study was to examine the development of children's inventions of vocal song endings over the course of three years. Children ($N = 52$) were followed for three years beginning in first-grade. They were divided into three groups (i.e., music – El Sistema-inspired program, sports, control) based on their extracurricular activities. For three consecutive years, children were asked to invent song endings to given song beginnings based on a task from the ATBSS. They also sang "Happy Birthday", engaged in pitch matching tasks, and talked about music. Invented song endings were analyzed in terms of number of trials and creativity scores, and in relationship to singing skills, which were examined for accuracy and range. Musical and extra-musical qualities of all vocal creations were also analyzed, giving origin to in-depth case studies of song creators and their unique products. Quantitative analyses revealed no interactions between groups and creativity scores or number of trials in any given year. Significant differences in creativity scores were found for year 3, for all groups. Qualitative analyses suggested that there were six different strategies used by child creators when inventing song endings, with the latter being influenced by children's personality, dispositions, and immediate social environments. The act of inventing song endings may come naturally to some children, irrespective of music training. These results suggest a multitude of contributing factors to executing and being comfortable with improvisation, including the importance of offering ample opportunities for children to improvise and compose in formal music education settings.

Factors Influencing Young Children's Singing Development

Steven M. Demorest[1], Simone Dalla Bella[2]
[1]Bienen School of Music, Northwestern University, Evanston, IL, USA
[2]Euromov Laboratory, University of Montpellier, Montpellier, France

The development of singing accuracy in children, and the relative role of training versus maturation, is a central issue for both music educators and those within music cognition. Singing ability is fundamental to children's developing musicianship and one of the primary goals of early music education. Despite this focus, a number of people have difficulty singing accurately even into adulthood and many report feeling inadequate as singers despite having normal skills in other domains. Perceptions of inadequacy can deter both children and adults from participating in music education and music-related activities. Is inaccurate singing a cognitive deficit or the result of a lack of proper music education? This symposium presents three controlled studies of factors that may influence young children's singing development. The first experiment explored the impact of early childhood music instruction on the singing development of children before they enter school. The second experiment examined the impact of daily singing instruction on 5-6 year olds' singing development in their first year of schooling. The third comparative study explored the role of instrumental training in singing accuracy of children age 5-12 years old. Taken together, this research can help to provide a picture of how singing development may be facilitated by certain kinds of musical experience and how those experiences interact with age and gender. Understanding how children's singing can be improved my shed light on some of the root causes of inaccurate singing and allow future generations to avoid some of the well-documented difficulties of adults who struggle with singing.

The impact of early childhood music education on singing abilities

Sean Hutchins[1,2]
[1]The Royal Conservatory of Music, Toronto, ON, Canada
[2]McMaster University, Hamilton, ON, Canada

Recent data show singing abilities can vary greatly in the adult population, and that several different factors can be involved in failure to sing in tune. However, the origins of singing ability (or difficulties) are poorly understood. Music educators often cite music education, especially in early childhood, as a key factor in developing this ability, but there is little in the way of acoustical data to support this assertion. In the present study, we aimed to test whether early childhood music education can improve singing ability, as assessed through acoustic measurements of intervallic accuracy. We tested two sets of children, aged 2.5 to 6 years. The first set (48 children) had taken weekly early childhood music education classes at The Royal Conservatory of Music for the past 24 weeks (as part of a larger, controlled study), the other set (61 children) had applied to take the same classes, but had been assigned to a control group (to receive the classes at a later time). In our testing session, children sang the alphabet song, in addition to several other musical and cognitive tasks. Acoustic analysis of their intervallic accuracy showed a significant interaction between Test Group (whether they had taken the lessons or not) and Gender, such that females in the lessons group made fewer interval errors that in those in the no lessons group. We also found improvement in accuracy across increasing age groups, which also interacted with gender. Singing ability also correlated with other factors as well. These results show some initial evidence that early childhood music education can have an impact on some singing abilities, which may be a precursor to better developed general musical skills throughout life.

The effect of focused instruction on young children's singing accuracy

Steven M. Demorest[1], Bryan Nichols[2], and Peter Pfordresher[3]
[1]Bienen School of Music, Northwestern University, Evanston, IL, USA
[2]School of Music, The University of Akron, Akron, OH, USA
[3]Department of Psychology, University at Buffalo, the State University of New York, Buffalo, NY, USA

Background & Method
Singing is one of the most natural ways for a child to engage in making music and singing activities play a central role in elementary music curricula. While singing ability seems to improve naturally for many children, there are those who struggle to match pitch and are less inclined to participate in singing experiences. Are there approaches to early childhood music education that might promote the acquisition of singing skills and help children avoid these challenges later in life?

We explored the effect of daily group singing instruction versus no formal instruction on the singing accuracy of kindergarten-aged children and whether singing accuracy performance differed across tasks. The treatment group (n = 41) received 20 minutes a day of group instruction in a Kodaly-based music classroom that emphasized the development of the singing voice in terms of tone, register, and accuracy for six months. The control group (n = 38) received no formal music instruction during the school day. Singing accuracy was measured both pre and post instruction by having participants echo single pitches, intervals and patterns sung by a recorded female non-vibrato vocal model and then sing Twinkle, Twinkle Little Star on their own.

Results & Conclusion
We found that all students showed improvement on the three echo-singing tasks from pre to post, but the experimental group demonstrated significantly more improvement. Performance on the familiar song task did not improve for either group. Students achieved the highest accuracy scores on the interval-matching task. Regular singing instruction seems to accelerate the development of accurate singing for young children, but the improvement is not evident in every singing task. Accuracy on shorter echo-singing patterns seems to precede the development of accuracy in singing songs despite the fact that singing songs is a much more commonplace activity for young children.

The impact of instrumental tuition on the development of children's singing competency

J.A. Saunders[1], G.F. Welch[1], E. Himonides[1], M. Sarazin[2] & V. Moulton[1]
[1]International Music Education Research Centre (iMerc), University College London Institute of Education, London, UK
[2]Wolfson College, University of Oxford, UK

The paper presents findings from *In Harmony* instrumental programmes that provide sustained instrumental tuition for children (aged 5-12 years) in two Primary schools in the UK. As a government-funded programme, the *In Harmony* initiative in England is based on the *El Sistema* orchestral social and music programme in Venezuela and, as such, are located in urban schools characterised by challenging circumstances. The aim of the research evaluation was to assess particular features of musical behaviours and development, children's attitudes to musical learning, and possible wider impact. Longitudinal data were collected via (a) specially designed questionnaire instruments for children that drew on different aspects of their experiences in music and of school life in general, (b) instrumental tutor reports on individual children's instrumental learning and (c) individual singing assessments. During visits made to each school, a researcher listened to each child's singing of two target songs and made a judgement as to the level of competency displayed against developmental criteria of two rating scales. The researcher's ratings were combined and converted into a 'normalised singing score' (NSS). Subsequent analyses revealed that the NSS values ranged from 22.5 (chant-like singing) to a maximum of 100 (no obvious errors in the musical features of the songs). Comparisons of children's NSS values were compared over time. Results indicate that whilst the main focus of the intervention was to support the development of instrumental skills, pupils entering the programme with significantly less developed singing competency (compared to children of a similar age and background in a national dataset of 11,000+ children) had made highly significant progress in their singing competency and were more in-line with children of the same age in the national dataset. This improvement was evidenced for both girls and boys, although sex differences were evidenced (in line with national and historical data) with girls being more developed than boys. The described measure of children's singing represents one indicator of their musical behaviour and development. We hypothesize that the finding of significant improvements in children's singing are likely to be, in part, related to the additional singing opportunities provided within the school week, but also to be an outcome of the auditory imagery or 'inner hearing' that occurs as a form of silent singing whilst children play their instruments, including when they play from musical notation.

Developing orchestration theory from an empirical analysis method

Meghan Goodchild[1], Stephen McAdams[1]
[1]Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT), Schulich School of Music, McGill University, Montreal, QC, CANADA

Orchestration theory has not attained the depth or precision compared to other musical fields and would benefit from perceptual grounding. Analyses of scores and treatises reveal that many orchestration aims are related to auditory grouping principles, which determine auditory fusion of musical events, the integration of coherent streams of events, and chunking into perceptual units. These grouping processes are implicated in orchestral devices related to timbral blend or heterogeneity, segregation or stratification of layers, and orchestral contrasts, such as antiphonal exchanges. We developed an empirical analysis method to study orchestral devices in a corpus of orchestral pieces. Pairs of researchers analyzed each movement: after analyzing the movement individually, they compared analyses and developed a joint analysis, which consisted of an annotated score and a spreadsheet that catalogued various details about each instance of the orchestral devices. These data include the span of measures in which the device is found, the timing of the recording(s), the instrumentation, and the perceived strength of the effect. A relational database was created, recording over 4200 orchestral devices from over 75 orchestral movements drawn from the classical period to the 20th century. Text data mining techniques of association analysis were used to explore the catalogued data in order to reveal patterns among instruments and their combinations within various orchestral devices. This approach will quantify the extent to which instrument combinations described in textbooks occur and also potentially uncover unknown combinations or unique instances by certain composers. Further analyses will investigate whether instrument combinations vary over time, as well as the contexts in which the devices occur. The aims are to contribute to the development of a theory of orchestration based on empirical analysis and to create a catalogue of excerpts for perceptual experimentation.

Factors influencing instrument blend in orchestral excerpts

Stephen McAdams[1], Pierre Gianferrara[2], Kit Soden[1], Meghan Goodchild[1]
[1]Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT), Schulich School of Music, McGill University, Montreal, QC, CANADA
[2]Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT), Cognitive Science Program, McGill University, Montreal, QC, CANADA

Most research on the blending of musical timbres has studied isolated sounds, rather than musical excerpts. We sought to understand the relative roles of the number of instruments, the diversity of their combinations within and across instrument families, the intervals employed, pitch register, and the degree of parallel or similar motion on blend perception in orchestral music. A sample of 64 excerpts representing different combinations of the above factors were selected from the analysis of over 70 movements of the orchestral repertory from the early classical period to the early 20th century. The excerpts were also chosen to cover a range of strengths from weak to strong blend. High-quality interpretations of these excerpts were created to mimic specific commercial recordings using the OrchSim system developed at McGill. From these interpretations, only the groups of instruments involved in a given blend were extracted and rendered to stereo sound files, thus initially testing the blending instruments outside of their surrounding musical context, but balanced within that context. Listeners rated the degree of blend of each excerpt on a scale from unity to multiplicity, with the scale ends representing the strongest blend and no blend, respectively. A regression analysis indicates that the primary factor affecting blend is the degree of similar motion. The effect of the number of instruments is more variable for smaller sets of instruments, where the timbral characteristics of the combined instruments may play a stronger role in determining blend. The long-term aim is to develop a model of blend perception that can be used for music information research and for computer-aided orchestration systems, as well for pedagogical applications in music theory, composition, and performance technique.

Contradictions in the Perception of Organ Stop *Vox humana*: From a Bear's Roar to the Human Voice. An Analysis and Comparison of Verbal Meaning and Sound Production

Girėnas Povilionis,[*1] Rima Povilionienė,[#2]

[*] *Centre for the Organ Heritage, Lithuania*
[#] *International Semiotics Institute, Kaunas University of Technology, Lithuania*
[1] girenas@vargonai.com, [2] rima.povilioniene@ktu.lt

ABSTRACT

The verbal naming of certain music timbres is considered to be a highly inventive aspect in the terminology of pipe organ stops. Most titles match the produced sound of stops, e.g. *Flute*, *Violina*, *Gamba*, *Posaune*, *Trompete*. Though some eloquent expressions (such as *Vox humana*, *Vox Angelica*, *Unda maris*), their etymology and attribution to certain organ sound present a contradictory centuries-long process. The article focuses on *Vox humana* (Lat. 'human voice'), one of the oldest reed stops with short resonators, of the Regal type. Its subtle manufacture and voicing is proof of an organ builder's excellence; there have been distinguished over 10 shapes of complex pipe scales that differ even in the frame of the same organ building school or master. The history of stop naming displays an intriguing conflict of verbal and sound perception. The origin of the title goes back to the beginning of 16th c. France, where *Voix Humaine* was used alongside *Jeu d'enfants*. Eventually organ builders in European countries adapted their versions: *Voz Humana*[-*e*] (used in Spanish organs), *Voce Umana* (Italy), *Anthropoglossa*, *Menschenstimme* or witty *Bärpfeife* ('bear pipe') in Germany. The article highlights the disagreement between the title and sound of the stop based on the excerpts from treatises written from the 17th to beginning of the 20th centuries. A systematization and comparison of pipe construction types that were manufactured in the 17–19th c. are presented.

INTRODUCTION

Regarded as one of the oldest baroque pipe organ stops, *Vox humana* stands out among the most refined reed stops, with short resonators of a cylindrical and/or conical form and of the Regal type. Usually this stop, being second to the more frequent reed type *Trompete*, was specific for large baroque organs with two or more manual keyboards. However, it is noted, that some one-manual organs in Sweden (dated end of the 18th and beginning of the 19th c.) or Lithuania (dated middle of the 19th c.) featured *Vox humana* beside *Trompete*.

For over 500 years of its existence *Vox humana* experienced plenty modifications of pipe scales, changes of the title and therefore accumulated a variety of sound-timbre features. There are over 10 different resonator shapes of *Vox humana* pipe scales; they differ among the examples attributed to the same organ building school and even produced by the same organ master. The manufacture of stop pipes requires a complex and thorough calculation, accuracy of the master and a long lasting process of voicing, and is widely considered as proof of the organ builder's excellence. So could it have been an organ master's dream to produce an organ stop that could perpetuate itself and convey personal voice – *Vox factorem*? This article focuses on mostly baroque types and the variety of *Vox humana* stops, its spread among countries and different organ building traditions.

VOX HUMANA IN EUROPEAN ORGAN TRADITION

The continuous implementation of the stop, titled with different versions of the Latin *Vox humana* ('human voice'), is noticeable in late renaissance and baroque pipe organs by various European organ building traditions starting from the middle of the 16th c. till the middle of the 19th c., when the romanticized transformations emerged. Mostly the stop was typical of instruments of the French and German organ building style, and directly related ones. However, according to Wedgwood, some ancient German organs featured *Pilgelchor* equipment – a *Vox humana* like effect "to represent the distant singing of pilgrims" (Wedgwood, 1905, p. 118).

Some of the origins of using the *Vox humana* stop in the organs go back to the middle of the 16th c. – French pipe organs, the stop besides *Voix Humaine* was called *Jeu d'enfants* (Lat. 'voice of a child') as well. In c. 1600 *Vox humana* was mentioned in Belgium, the Netherlands (e.g., master Neijenhoff, c. 1600, Schiedam St. John's Church, The Netherlands), and later – in Spain (*Voz humana[e]*, e.g., unknown master, 1762, Tarifa St. Matthew's Church, Spain) (see Eberlein, 2009; Stauff, 2016). Through the Bader family of organ builders that operated in the Netherlands and Westphalia, and the organ master Arp Schnitger the use of *Vox humana* was introduced to Westphalien and North German territories in the middle and second half of the 17th c. At the same time there are some known examples in English pipe organs. Since 1700 the stop was introduced to the East and Middle Germany (through the masters Claude Legros and Andreas Silbermann). In c. 1730 through the organ masters Johan Niclas Cahman (c. 1670–1737) and Lambert Daniel Kastens (c. 1690–1744) the *Vox humana* stop was brought to Scandinavian countries. There, e.g. in Sweden, organs by Wahlberg used the reed stops *Trompete* and *Vox humana*, and by the way, *Vox Virginea* (1764 organ in Karlskrona Fredrikskyrka). And in c. 1750 the Badische organ masters took over the stop from the master Johann Andreas Silbermann (1712–1783).

The German tradition also used another version of the title – *Anthropoglossa*, *Menschenstimme* or even the witty *Bärpfeife* ('bear pipe'). The use of the latter still marks some uncertainty. E.g., Werckmeister's remark that *Bärpfeife* sometimes dubbed *Vox humana* is disputed by Wedgwood pointing to it as "truly a questionable enough compliment to the human voice" (Wedgwood, 1905, p. 6); on the other hand, Audsley notes, that *Bärpfeife* "of 2 ft. and 1 ft. have whistling tones resembling those of the human mouth" (Audsley, 1925, p. 41).

The spread of *Vox humana* in Lithuanian baroque organs, starting the 2nd half of 18th c., was affected by the East

Prussian organ building school of German legacy by the master Adam Gottlob Casparini (1715–1788), who settled in Königsberg, and possibly – Gerhardt Arendt Zelle (?–1761), the founder of Vilnius Late Baroque Organ Building School. Thereafter, from the 2nd half of the 18th c. to the middle of the 19th c., *Vox humana* remained one of the main exclusive reed stops in almost all large (2-keyboard manual) Lithuanian baroque organs as well as portatives (small organs).

Italian baroque organs featured the same title *Voce Umana* as well. However, in the organ heritage of this country we encounter a different type of the stop and its pipes. Italian *Voce Umana* is characterized by labial pipes instead of reed, and every tone is produced by two unevenly tempered pipes that produce a wavy sound (e.g., organs by the Antegnati family masters in Bergamo St. Nicola Church, 1588, and Brescia St. Carlo Church, 1630, Italy).

TIMBRAL AND CONSTRUCTIVE PECULIARITIES

Usually an 8-foot *Vox humana* reed stop was produced seeking to imitate the human voice. However, an English music historian Charles Burney (1726–1814), who had listened to the sound of many glorious organs in Europe (e.g., Gottfried Silbermann's 1736-organ in Frauenkirche in Dresden) and set down his tour impressions in the famous volume *The Present State of Music in Germany, the Netherlands, and United Provinces* (1st ed. 1773), stated after a disappointing visit to Müller's 1738-organ in Haarlem church: "[...] the world is very apt to be imposed upon by names. [...] I must confess that, of all the stops I have yet heard which have been honoured by the appellation of Vox Humana, no one in the treble part has ever yet reminded me of anything human so much as of the cracked voice of an old woman of ninety, or in the lowest parts of Punch singing through a comb." (Burney, 1775, p. 305) Though in some cases Burney praised this stop as "one of the best [...] of that kind" (e.g., Amsterdam church organ) and expressed modest compliments on the "sweet" *Vox humana* like sound of a fine oboe or clarinet, but, still, not "a human voice". Later on James I. Wedgwood presented with Robertson's note, recorded in 1897, that the thin and nasal *Vox humana* tone may resemble "anything, from Punch's squeak to the bleating of a nanny goat", adding his own remarks "a ludicrous caricature of the human voice" or "the peculiar 'flavour' of the stop has also led to the mock-name of the 'gas-pipe'" (Wedgwood, 1905, p. 178–179).

There are not so many surviving authentic examples of centuries-old *Vox humana* worldwide; most baroque organs were rebuilt in the 19th c., usually *Vox humana* was removed or replaced during the changes of the stop list. However, the comparison of known examples shows a very different approach to and manufacture of the same titled stop.

It is worth noting that in Lithuania, a remote East-Central European country, the authentic baroque *Vox humana* has survived in several organs: in Kretinga Franciscan church (unknown master, end of the 17th c.), Vilnius church of the Holy Spirit (Prussian master Adam Gottlob Casparini, 1776), Tytuvėnai church (Vilnius school master Nicolaus Jantzon, 1789), Žemalė church (unknown master, 1839), and some fragments in Kurtuvėnai church (Vilnius school master Mateusz Raczkowki, 1792–3).

Nowadays, some examples of *Vox humana* pipe construction are distinguished as the most representative and typical examples among others. Seeking to produce the delicate and rapid pulsation of sound like the vibrato effect, *Vox humana* most commonly was performed in a combination with other 8-foot registers with stopped pipes or *Tremulant* (in some cases the latter equipment was installed inside the *Vox humana* as a double-stop construction) (Friedrich, 1989, p. 45).

The *Vox humana* pipe consists of two main parts: a metal or wooden boot (bottom part) and metal (very rarely wooden) resonator (top part). The quality of timbre and sound especially depends on the construction and shape of resonator that usually was produced using two main geometrical forms of cylinder and conic (plus a conic with flat top) and various combinations (see Fig. 1)

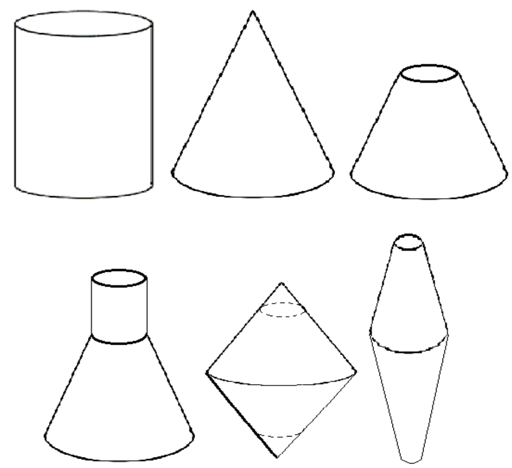


Figure 1. The geometrical forms as the basis of *Vox humana* pipe construction: cylinder, conic, conic with flat top, and their combinations: cylinder and conic with flat top, double conic (and/or with flat top), spindle.

Types of Construction of *Vox Humana*

The historically settled types of *Vox humana* pipe construction and attributed examples are:

1. Half-stopped resonator with cylindrical top section, bottom section – inversion of cone with flat top. Example: *Voix humaine* in Meaux Cathedral (France), 1627, master Valéran-de-Héman (Bédos, 1766, engraving No. 18, fig. 141).

In 17–18th c. France, this stop experienced slight modifications, the resonators of lower tones were prolonged. Similarly *Vox humana* was produced in the 17th c. Netherlands (example: organ by the Duyschot family masters, c. 1687, Leiden St. Peter Church).

In c. 1670 a *Vox humana* with a cylindrical resonator appeared in the Adlington Hall organ, Macclesfield (UK). This type of resonator may be linked to the organ master Bernard Smith and was adopted by German masters, e.g., Andreas Silbermann, Gottfried Silbermnann and Johann Michael Stumm. Examples of type 1 presented in Fig. 2.

This type of construction is quite similar to the resonator and pipe scale of *Zink*[*e*] (also *Zinck*[*e*], *Cink*[*e*]) stop. The latter example has survived in Kretinga Franciscan Church (unknown master, end of the 17th c.; Lithuania), its inauthentic title *Trompet* 8' was made in the 19th c. (see Fig. 2 & 3).

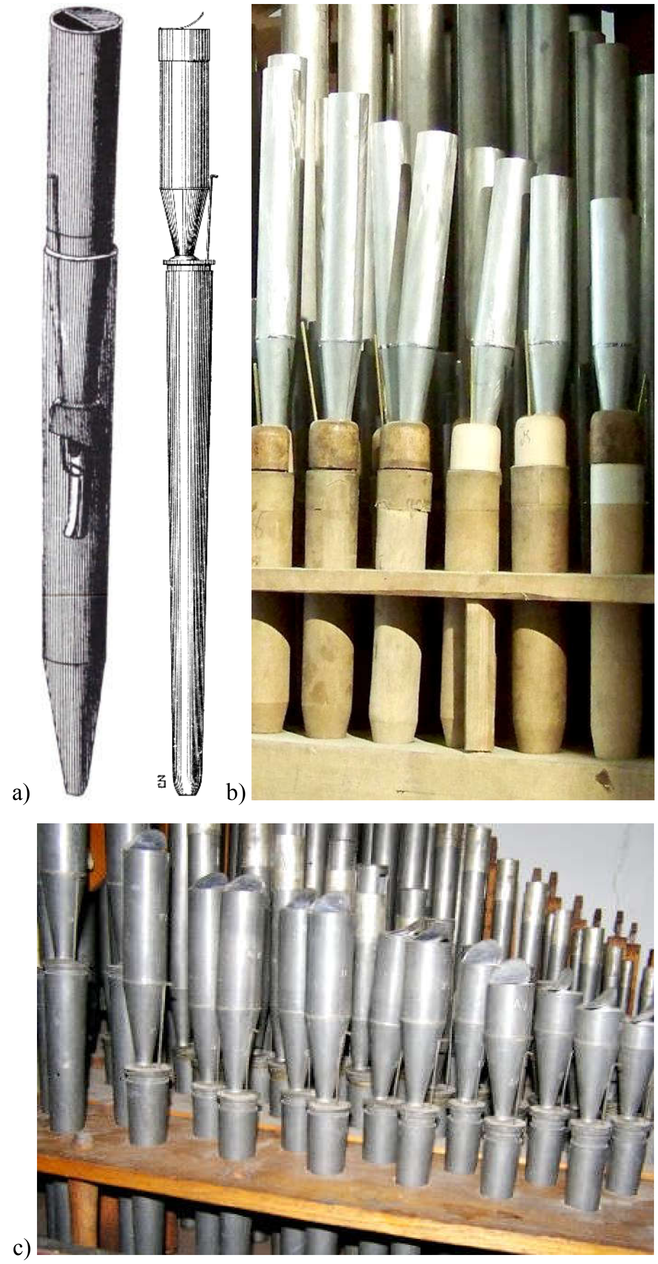


Figure 2. Pipe examples of Type 1: a) *Vox humana* drawings (left – Bédos, 1766, engraving No. 18, fig. 141; right – Audsley, 1905, fig. CCCLVI); b) *Trompet* pipes in Kretinga Franciscan Church (unknown master, end of the 17th c.; Lithuania); c) newly manufactured (2008) *Vox humana* pipes of this type for the organ in St. Paul's Stoneycroft, Liverpool (UK).

2. Double conical resonator on cylindrical foot, wide cylindrical resonance hood with central cavity and small surrounding holes. This type was exploited by the master Arp Schnitger in the 2nd half of the 17th c. This type of pipe has survived in the organ by Dietrich Christoph Gloger, 1758–63, Cadenberge St. Nicolaus Church (Germany). Albertus Antonius Hinsz continued Schnitger's organ building tradition in the Netherlands. The manufacture of his *Vox humana* in Kampen St. Nicholas Church (The Netherlands, 1741–3) is an obvious example of Schnitger's influence (see Fig. 3).

Like the first type, the second type is quite similar to the resonator and pipe scale of *Zink*[*e*] (also *Zinck*[*e*], *Cink*[*e*]) stop (see Fig. 3).

Figure 3. Pipe examples of Type 2: a) the copy of the *Vox humana* pipes from Arp Schnitger's organ in Hamburg Hauptkirche St. Jacobi (1688–93; a copy of Schnitger's organ installed in new organ in Gothenburg Örgryte New Church and implemented by GOArt/Göteborg Organ Art Centre, Sweden, in 2000); b) *Vox humana* by Albertus Antonius Hinsz (1741–3, Kampen St. Nicholas Church, The Netherlands); c) an example of analogous construction – the copy of the *Zink* stop from Arp Schnitger's organ in Hamburg Hauptkirche St. Jacobi.

3. Double conical resonator on cylindrical foot, without hood, open. Known examples: Gottlieb Heinrich Herbst, 1728–32, Schlosskirche in Lahm, Itzgrund (Germany). Similar pipes specific for the organ by Christian Müller, 1738, Haarlem the Grote Kerk or St. Bavokerk (The Netherlands); by the Paris master Jean-François Lépine, 1759, Béziers St. Nazaire Cathedral (France; the stop titled *Voix humaine a l'allemande* 8'). This type of newly manufactured *Vox humana* pipes was used for the old organ by Joseph Gabler (1739–50, Weingarten Basilica of St. Martin, Germany, see Fig. 4).

This type of stop was called *Basson ou Voix humaine a l'allemande* 8' as well; in Bédos' catalogue the stop described as *Basson* (see Bédos, 1766, engraving No. 129, fig. 4). By the way, the Bédos' catalogue provides a quadruple type of conical-shaped resonator (see Fig. 4).

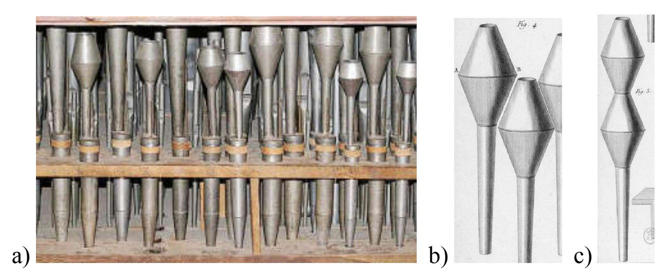
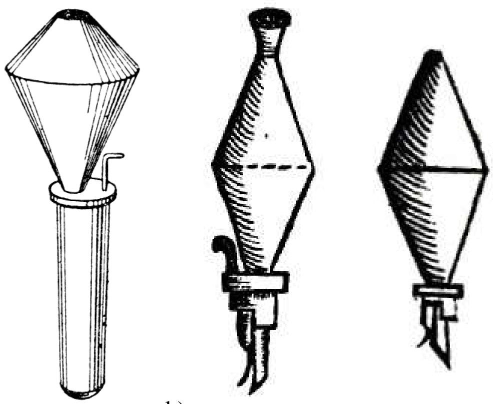


Figure 4. Pipe examples of Type 3: a) newly manufactured *Vox humana* pipes in the old organ by Joseph Gabler (1739–50, Weingarten Basilica of St. Martin, Germany); b) a drawing in Bédos, 1766 (engraving No. 129, fig. 4); c) an example of quadruple combination (Bédos, 1766, engraving No. 129, fig. 5).

4. Double conical resonator, open. The type is analogous to the *Bärpfeife* stop (see Fig. 5). It is known that the organ master Arp Schnitger produced this type of pipe (with asymmetric double cones, lower part longer than top) for the 1699 organ in Uithuizen Church (The Netherlands).

This type of resonator was used by East Prussian organ builders, and was discovered in the organ by Andreas Hildebdrandt, 1746–7, Gdansk St. Barbara's Church (Poland). A similar example is specific to Zaharias Hildebrandt's organ, 1743–6, Naumburg St. Wenzel Church (Germany). Hinsz used this type for some *Vox humana* pipes in his 1741–3 organ in Kampen St. Nicholas Church besides other pipes of type 1 construction (with a wide cylindrical resonance hood, see Fig. 5).



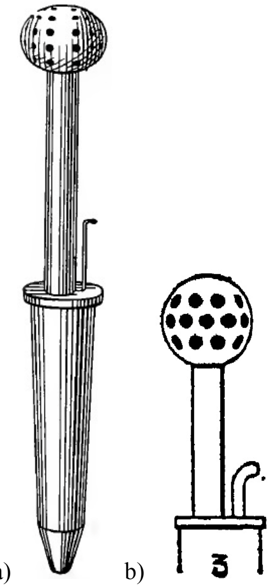
a) b)

c)

Figure 5. Pipe examples of Type 4: a) a drawing of a traditional *Bärpfeife* pipe (author unknown); b) a drawing of *Bärpfeife* (left) and *Vox humana* (right) pipes by Praetorius (Praetorius, 1619, fig. XXXVIII); c) *Vox humana* pipe with double conical resonator by Albertus Antonius Hinsz (1741–3, Kampen St. Nicholas Church, The Netherlands).

5. Resonator with cylinder bottom section and spherical (apple-shaped) head, pierced with small holes (Germ. Apfelregal, Knopfregal). Possibly, it is one of the oldest types of organ pipe construction but there are no surviving examples. Among mentioned historical examples are: the master Christoph Contius, 1713–6, Halle Marktkirche; Johann Georg Schröter, 1716, Erfurt Augustine Church (Germany). In his dictionary of organ stops, Wedgewood provides a drawing of the so-called *Apfel-Regal* pipe (see Fig. 6).

6. Two-section tapered resonator, with dome-shaped top (hood) and round hole on the side (see Fig. 7). A historically mentioned, but not survived example is: Joseph Gabler, 1739–50, Weingarten Basilica of St. Martin (Germany).



a) b)

Figure 6. Pipe example of Type 5: a) a drawing of *Apfel-Regal* pipe by James Wedgewood (Wedgewood, 1905, p. 134); b) a drawing of *Apfel-Regal* pipe by George Audsley (Audsley, 1921, p. 224)

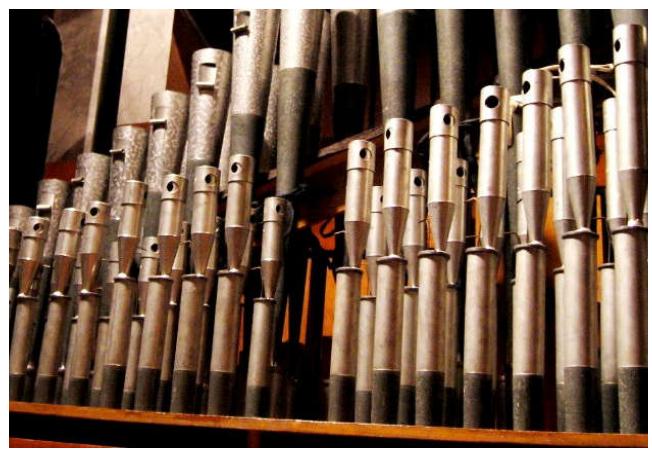


Figure 7. A newly manufactured pipe example of Type 6 by Robert Morton Organ Co. (1928, Ohio Theatre, Columbus, USA).

7. One-section tapered resonator, half-stopped. A historically mentioned example is: Johann Gottlieb Benjamin Engler, 1813–22, Wrocław St. Mary Magdalene Church (Poland); the organ featured a very rare example of *Vox humana* because of wooden square-tapered resonators from tone *c*.

8. One-section tapered resonator, stopped, with vowel cavity on the side of the top. A historically mentioned example is: Johann Daniel Boden, 1776, Sambleben in Schöppenstedt, Germany).

9. Spindle-shaped resonator (the length of top conic varies from equal parts to 2-3 times shorter due to the pitch), with vowel cavity on the top. Examples: Andreas Hildebrandt, 1719, Pasłęk St. Bartholomew Church (Poland); Adam Gottlob Casparini, 1776, Vilnius Church of the Holy Spirit (reconstructed in 2006); Nicolaus Jantzon, 1789, Tytuvėnai church; Mateusz Raczkowski, 1792–3, Kurtuvėnai Church; unknown master, 1839, Žemalė Church (all mentioned in Lithuania) (see Fig. 8).

a) 

b) 

c)  d) 

Figure 8. Pipe examples of Type 9: a) by Andreas Hildebrandt (1719, Pasłęk St. Bartholomew Church, Poland); b) by Adam Gottlob Casparini (1776, Vilnius Church of the Holy Spirit, Lithuania); c) by Mateusz Raczkowski (1792–3, Kurtuvėnai Church, Lithuania); d) by unknown master (1839, Žemalė Church, Lithuania).

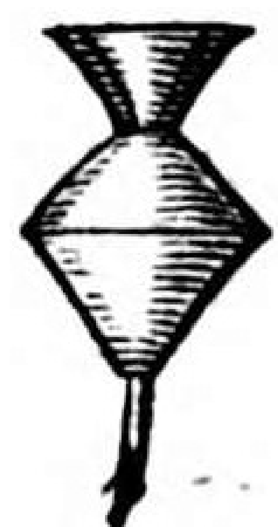
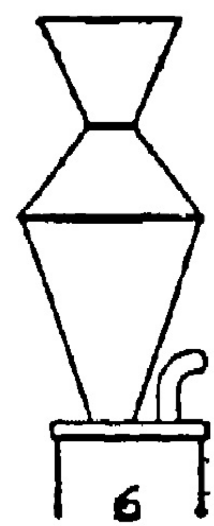
Types of *Vox Humana* Construction in the 19th–20th C.

1. Triple conical resonator. This type of pipe construction was more common in the pipe organs of romantic style, starting the 19th c. However, it is analogous to the old *Bärpfeife* stop construction; its triple construction was already mentioned by Adlung in 1726 (his book published in 1768) and later authors – Seidel (1844), Wedgewood (1905), Audsley (1921), etc.

In c. 1829 the organ master Ignaz Bruder in his workshop records book mentioned producing *Vox humana* with triple conical shape resonators. An example of *Vox humana* pipes of the famous Bätz-organ in the Hague (Fig. 9) was possibly manufactured during the rebuilding of the organ (the baroque organ was built in 1762 by Johann Heinrich Hartmann Bätz, but in 1837 his grandson Jonathan Bätz made the essential changes in the stoplist).

In the 20th c. this type was manufactured by the organ building workshop of Klais Orgelbau for Würzburg Cathedral (Germany, 1968–9).

a) 

b) c)

Figure 9. Example of triple conical resonator: a pipe from the organ by organ master Johann Heinrich Hartmann Bätz (1762) and rebuilt by Jonathan Bätz (1837, Evangelical Church, The Hague, The Netherlands); b) drawing of the *Bärpfeife* of triple-conical construction (Adlung, 1768, p. 73); c) drawing of the *Bärpfeife* of triple-conical construction (Audsley, 1921, p. 224).

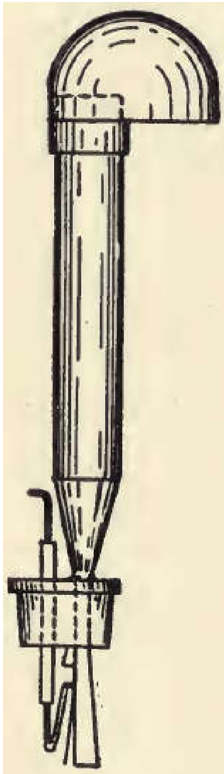


Figure 10. Example of conical-cylindrical-spherical combination: a drawing of *Vox humana* pipe in Norman and Beard's organ in Norwich Cathedral (1899, England) (Miller, 1913, p. 92, fig. 20).

2. One-section tapered resonator, with closed cover. Example: organ by Jemlich workshop, 1899, Lößnitz Johanniskirche (Germany), the stop is placed in a distant box over the organ case. This type of newly manufactured pipe was introduced in 1932 and 1956 – catalogues by 20th c.-organ building workshop of August Laukhuff.

3. Free reeds (Germ. *Durchschlagende Zungen*). The first known examples are dated the end of the 18th c.; in the 19th c. it was the most popular type of resonators among organ building workshops, e.g., organ by the Walker workshop, 1879, Düsseldorf Johanneskirche (Germany).

4. A combination of conical bottom section, cylindrical middle section and half-spherical resonance hood. In 1913, an interesting and rare example of *Vox humana* pipe construction was printed in Miller's study about organ building, while discussing the organs of the 19th c., and referring to Norman and Beard's organ in Norwich Cathedral (Miller, 1913, p. 92; see Fig. 10).

Labial Type of *Vox Humana*

This kind of organ stop features the vibrant, wavy (Germ. *Schwebende*) and slightly untuned sound (uneven tuning), and is attributed to the group of Principal stops. To this day the stop has been used in Italian organs since c. 1550, with the title *Voce umana* or *Fiffaro*, *Piffara*, *Bifara*. Furthermore, this type of stop was brought into the surroundings of Germany and later spread more widely through the organ builder Eugenio Casparini (Johann Caspar, 1623–1706), who after returning from Italy, manufactured this stop for the 1697–1703 organ in Görlitz St. Peter and Paul Church. This type is quite analogous to another labial stop *Unda maris*, which is more common.

ACKNOWLEDGMENT

The preparation of the article and participation at ICMPC14 partly supported by the Lithuanian Composers' Union and Kaunas University of Technology.

REFERENCES

Adlung, J. (1768). *Musica Mechanica Organoedi*. Teil 1. Berlin: Friedrich Wilhelm Birnstiel.

Audsley, G. A. (1905). *The Art of Organ Building*. Dodd, Mead & Company.

Audsley, G. A. (1921). *Organ-Stops and their Artistic Registration. Names, Forms, Construction, Tonalities, and Offices in Scientific Combination*. New York.

Bédos de Celles, D. F. (1766). *L'Art du Facteur d'Orgues.* Paris; faximile ed. (1994): Geneve.

Burney, C. (1775). *The Present State of Music in Germany, the Netherlands, and United Provinces: Or, the Journal of a Tour Through Those Countries, Undertaken to Collect Materials for a General History of Music*, 2nd ed., vol. 2. London.

Eberlein, R. (2009). *Orgelregister, ihre Namen und ihre Geschichte*. Siebenquart: Verlag Dr. Roland Eberlein.

Friedrich, F. (1989). *Der Orgelbauer Heinrich Gottfried Trost: Leben-Werks-Leistung*. Leipzig:Deutscher Verlag für Musik.

Miller, G. L. (1913). *The Recent Revolution in Organ Building*. 2nd ed. New York: The Charles Francis Press.

Praetorius, M. (1619). *Syntagma Musicum.* 2nd book 'De Organographia'. Wolfenbüttel.

Praetorius, M. & Faulkner, Q. (trans. & ed.) (2014). *Syntagma Musicum II: De Organographia*, Parts III – V with Index. Lincoln: Zea E-Books. Book 24.

Seidel, J. J. (1844). *Die Orgel und ihr Bau*. Breslau.

Schneider, T. (1958). *Die Namen der Orgelregister. Kompendium von Registerbezeichnungen aus alter und neuer Zeit mit Hinweisen auf die Entstehung der Namen und deren Bedeutung*. Kassel, London: Bärenreiter.

Smets, P. (1948). *Die Orgelregister, ihr Klang und Gebrauch: ein Handbuch für Organisten, Orgelbauer und Orgelfreunde*. Mainz: Rheingold-Verlag.

Stauff, E. L. (2016, April 28). Vox humana. *Encyclopedia of Organ Stops*. Retrieved April 28, 2016, from http://www.organstops.org/v/VoxHumana.html.

Wedgwood, J. I. (1905). *A Comprehensive Dictionary of Organ Stops (English and Foreign, Ancient and Modern): Practical, Theoretical, Historical, Esthetic, Etymological, Phonetic*, 3rd ed. London: The Vincent Music Company.

The effect of timbre differences on pitch interval identification in musically trained listeners

Sarah Gates[1,2], Stephen McAdams[1], and Bennett K. Smith[1]
[1]Schulich School of Music, McGill University, Montreal, Quebec, Canada
[2]now at Bienen School of Music, Northwestern University, Evanston, Illinois, USA

Changes in timbre (namely spectral centroid) can affect subjective judgments of the size of pitch intervals. We investigated the effect of timbre change on musicians' ability to identify melodic intervals in order to observe whether poorly identified pitch intervals are more susceptible to interference with timbre. Musicians (n=39) identified all melodic intervals within the octave in ascending and descending directions using traditional labels (e.g., "perfect fifth") in two timbre conditions: (1) timbre-neutral (piano only) and (2) timbre-change (3 timbre pairs). The pairs were chosen from a timbre dissimilarity experiment (n=19) using multidimensional scaling in order to determine the perceptual dimensions along which timbre change occurred. The pairs chosen were marimba—violin (spectral centroid), French horn—piano (temporal envelope), and bass clarinet—muted trumpet (amplitude modulation). In a speeded interval-identification task, musicians' vocal identifications of melodic intervals were analyzed in terms of accuracy and response time. A repeated-measures ANOVA showed a main effect of timbre on accuracy scores, indicating that the marimba—violin pair had the lowest scores overall. Timbre pair also affected the type of errors in incorrect trials. Root-mean-squared errors calculated from confusion matrices revealed an increase in the spread of errors for timbre-change conditions (the largest being the marimba—violin pair) compared to the timbre-neutral condition. The response-time analysis showed a significant main effect of timbre demonstrating that both the fastest (French horn—piano) and slowest (marimba—violin) response times were within the timbre-change condition. The results indicate that timbre-change along differing dimensions can affect the categorical identification of pitch intervals. The lack of systematic interaction between timbre pair, interval and direction suggests that this effect may not be limited to poorly identified melodic intervals.

Top-Down Modulation on the Perception and Categorization of Identical Pitch Contours in Speech and Melody

J. L. Weidema,[1] M. P. Roncaglia-Denissen,[2] H. Honing[3]

[1] *Music Cognition Group (MCG), Institute for Logic, Language and Computation (ILLC), Amsterdam Brain and Cognition (ABC), University of Amsterdam, The Netherlands*

[1]j.l.weidema@uva.nl, [2]m.p.roncaglia@uva.nl, [3]honing@uva.nl

I. Introduction

Both speech and music emphasize the acoustic signal. Pitch is a fundamental, and highly perceptual acoustic attribute of both domains. In speech, pitch is generally continuous and curvilinear, in music often relatively discrete [e.g., 1]. At present, there is no consensus on whether pitch in both domains is governed by distinct, domain-specific cognitive mechanisms [2,3] or whether it is processed by domain-general, shared processing mechanisms that span over both domains [4–6]. Contributing to consensus is further complicated when considering both the structural and functional differences between pitch in speech and music making direct comparisons between domains challenging [e.g., 6,7].

II. Aim

The aim of the present study is to investigate whether mechanisms governing pitch contour processing operate differently when pitch information is perceived as either speech or melody. It seeks to contribute to the question of domain-specificity by using analogous pitch contours as stimuli. By modulating listening mode, the present study aspires to demonstrate that contour processing relies on distinct domain-specific cognitive mechanisms that are regulated by top-down mechanisms from speech.

III. Method

Three groups of participants (Mandarin, Dutch, and musicians) were exposed to identical pitch contours in speech and melody. Critical items (CIs) for the speech condition consisted of disyllabic words in Mandarin with three possible contours: rising, falling, or flat. A pitch continuum was created for the CIs from rising to flat, and flat to falling, by dividing a finite pitch space into 11 equal-sized tonal gradations [see 8]. CIs for the melodic condition were exact analogues of the manipulated speech contours, created by extracting the pitch contour of the critical speech items, and resynthesizing these as gliding sinusoidal tones. Given condition, CIs were placed in linguistic, and musical carrier phrases, respectively. Participants indicated the direction of the last tone in a 3-alternative forced-choice task. Accuracy rates were collected.

IV. Results

Results indicate domain-specific perceptual mechanisms operant in both speech, and music. Identical pitch contours were perceived significantly different given listening mode, and group: While top-down influences from language appeared to significantly alter perception of pitch contour in speech for speakers of Mandarin, and Dutch differently, this effect was lacking in musicians. A general effect of pitch-expertise was found on the perception of melody.

V. Conclusion

The classification patterns of pitch contours in speech and music seem to suggest that domain-specific categorization is modulated by top-down influences from language and music. We provide further evidence that experience with pitch in one domain can exert significant influence on the manner in which pitch is perceived in another domain.

References

1. Zatorre RJ, Baum SR. Musical melody and speech intonation: singing a different tune. PLoS Biol. 2012;10: e1001372. doi:10.1371/journal.pbio.1001372

2. Peretz I, Coltheart M. Modularity of music processing. Nat Neurosci. 2003;6: 688–691. doi:10.1038/nn1083

3. Peretz I. Music, language and modularity framed in action. Psychol Belg. 2009;49: 157–175. doi:http://dx.doi.org/10.5334/pb-49-2-3-157

4. Asaridou SS, McQueen JM. Speech and music shape the listening brain: evidence for shared domain-general mechanisms. Front Psychol. 2013;4: 321. doi:10.3389/fpsyg.2013.00321

5. Patel AD. Music, language, and the brain. Oxford: Oxford University Press.; 2008.

6. Patel AD. Language, music, and the brain: A resource-sharing framework. In: Rebuschat P, Rohrmeier M, Hawkins J, Cross I, editors. Language and music as cognitive systems. Oxford: Oxford University Press; 2012. pp. 204–223.

7. Peretz I, Vuvan D, Armony JL. Neural overlap in processing music and speech. Phil Trans R Soc B 370 20140090. 2015; doi:http://dx.doi.org/10.1098/rstb.2014.0090

8. Xu Y, Gandour JT, Francis AL. Effects of language experience and stimulus complexity on the categorical perception of pitch direction. J Acoust Soc Am. 2006;120: 1063–1074. doi:10.1121/1.2213572

The structure of absolute pitch abilities and its relationship to musical sophistication

Suzanne Ross[1,2], Karen Chow[1], Kelly Jakubowski[1], Marcus T. Pearce[3], Daniel Müllensiefen[1]
[1]Department of Psychology, Goldsmiths, University of London, London, UK
[2]Centre for Music in the Brain, Department of Clinical Medicine, Aarhus University, Aarhus, Denmark
[3]School of Electronic Engineering and Computer Science, Queen Mary University of London, London, UK

Two types of absolute pitch abilities have been identified from previous research: overt AP (e.g. pitch labelling; oAP) is purported to be a rare binary ability possessed by a small proportion of people with a musical background, while latent AP (recognising or producing a well-known song at the correct pitch; lAP) is thought to exist in the general population and can be measured on a continuous scale. However, the measurement structure of these abilities (binary versus continuous) and the degree to which the two are related still needs to be confirmed. Furthermore, it may be that lAP is merely a side-effect of singing ability, musical engagement, or formal musical training. The relationship between lAP and musical sophistication thus requires clarification. We therefore developed of a comprehensive test battery for measuring oAP and lAP in musicians and non-musicians to address the aforementioned questions. 104 musician and non-musician participants were tested on five oAP and three lAP pitch production and perception tests, as well as three subscales of the Goldsmiths Musical Sophistication Index self-report inventory. In a preliminary analysis, Gaussian mixture modelling showed oAP scores to be bimodal and lAP to be unimodally distributed. Variable selection for cluster discrimination and exploratory factor analysis suggested different pitch production tests as the most efficient measures of latent oAP and lAP abilities. A point-biserial correlation indicated a relationship between overall oAP and lAP scores, but this relationship was not found when participants with and without oAP were analysed as separate groups. There was no significant correlation between lAP scores and active engagement, musical training or singing ability. These results support previous findings that oAP is a binary ability and indicate that lAP is a continuously expressed ability which is distinct from oAP. Results further show that lAP is not a mere side-effect of musical sophistication.

Heptachord Shift: A perceptually based approach to objectively tracking consecutive keys in works of J.S. Bach

Marianne Ploger[1]
[1]Blair School of Music, Vanderbilt University, USA

Widely acknowledged as the greatest master of tonal modulation, J.S. Bach's works elude traditionally triad-based analyses more applicable to later eras. Considerable disparities exist among theorists in the analyses of the same piece resulting in areas of tonal ambiguity. But what if there were a way to remove such ambiguities and clearly and objectively identify both by ear and eye the tonic and mode of each consecutive key throughout a piece by Bach? This lecture demonstration explores a cognitively based approach to modulation called *heptachord shift* in which the perception of the tonic and mode of each key is based on the mind's unconscious ability to bind together through time the most recently heard seven diatonic pitches to form an unbroken chain of objectively perceptible tonal heptachords from the beginning to the end of a piece of Bach, allowing us to know the specific degree 'meaning' of each note, and echoing how we comprehend language in real time— one word at a time. The Prelude in D major, BWV 850 acts as a model.

Perception of pitch accuracy in melodies: A categorical or continuous phenomenon?

Pauline Larrouy-Maestri[1], Simone Franz[1], and David Poeppel[1,2]
[1]Neuroscience Department, Max-Planck-Institute for Empirical Aesthetics, Frankfurt, Germany
[2]Department of Psychology, New York University, USA

In western music, a semitone constitutes a theoretical and perceptual boundary between tones and is defined as a unit for categorical perception of intervals (Burns & Ward, 1978). However, melodic perception does not rely exclusively on this category but also involves the notion of 'correctness'. If we usually classify melodies as "in tune" or "out of tune" depending on the size of interval deviations (smaller than a semitone) along melodies, the transition between the two categories remains unclear. This study examines the process involved in pitch accuracy perception.

Twenty-five participants identified melodies as "in tune" or "out of tune" and rated their confidence for each answer. The pitch manipulation consisted of the enlargement of an interval in 5 cent steps (from 0 to 50 cent deviation). The interval deviated was either a Major 2nd or a Perfect 4th and occurred in the middle or end of a 6-tone melody. The task was run twice, before and after an explicit definition of the two labels. Repeated measure ANOVAs were conducted to examine the effect of the deviation on the proportion of in-tune answers and on the confidence levels.

For the participants who were able to learn the labels (n = 20), the proportion of in tune answers varies greatly according to the amplitude of the deviation and depended on the size of the interval manipulated. Associated with the confidence level measurement, the identification data support a categorical perception process. Interestingly, explicitly learning the labels increased the overall confidence but did not modify drastically the profile of the categories and the process behind the categorization.

This study suggests that explicit learning is not necessary to develop higher order categories relative to "correctness". Nevertheless, such a process seems limited to certain intervals. Further investigation of other intervals and individual differences seems promising to better understand the mechanisms underlying music perception.

Isomorphism of Pitch and Time

Olivia Xin Wen,[*1] Carol Lynne Krumhansl [*2]
[*]*Psychology, Cornell University, USA*
[1]xw383@cornell.edu, [2]clk4@cornell.edu

ABSTRACT

An ongoing debate regarding the perception of pitch and time is whether information on the two dimensions is processed independently or interactively. To study this, we tested whether listeners prefer sequences in which tonally stable tones coincide with rhythmically stable tones. Our study builds on a noted isomorphism between pitch intervals in the diatonic scale and tone durations in the standard rhythm originating in Ghana. This isomorphism is shown in a) the maximally even structure of 2212221 and b) the cyclic nature with seven possible starting points. To better understand how pitch and time combine perceptually, we conducted two experiments. In Experiment 1, we created seven scales based on the diatonic pattern and seven rhythms based on the standard pattern by shifting the starting pitch interval or tone duration. To measure the perceived tonal stability of tones in the scales, in Experiment 1a each scale was followed by a probe tone and listeners judged how well the tone fit into the scale. To measure the perceived rhythmic stability of tones in the rhythms, in Experiment 1b each position of the sequences was accented dynamically and listeners judged how well the accent fit into the rhythm. These ratings were then used in analyzing the results of Experiment 2 that used all 49 pairs combining the 7 scales and 7 rhythms in Experiment 1. Participants rated a) how well the rhythm fits the scale for each pair and b) familiarity and well-formedness of each scale and rhythm. Results show that probe ratings from Experiment 1 predict judgments in Experiment 2. Specifically, higher ratings were given to scale/rhythm pairs in Experiment 2 when the tonal and rhythmic hierarchies found in Experiment 1 correlated more strongly with each other. In addition, we found a familiarity bias toward the major scale. After accounting for this bias, results remain significant. These suggest that information from the two individual dimensions interact perceptually.

I. INTRODUCTION

Pressing (1983) noted the isomorphism shown in Figure 1a between the pitch intervals in the diatonic scale in Western music and the durations in what is called the standard pattern. Whereas the diatonic scale is the most prominent scale in Western tonal music, the standard pattern is the most used African rhythmic pattern, originating in Ewe dance music in Ghana (Agawu, 2006). It can be found running in the background of Afro-Latin music, such as Salsa. The Standard Pattern is often played in a repeated fashion, meaning that, once the rhythm ends, it loops back to the beginning. This repetition can also be found in the domain of scales; it takes the form that the scale intervals repeat when the octave is reached. In fact, this cyclic feature is only one of the parallels between the diatonic scale and the standard pattern. As Figure 1b shows, this shared structure between scale and rhythm can be explained as a maximally even set (Clough & Douthett, 1991), where the seven white circles are maximally evenly spread among the twelve evenly distributed positions around a circle. On the circle, the distance between every two adjacent positions on the circle can represent a semitone in pitch or an eighth note in duration. Starting from the white circle at the 12 o'clock position and counting clockwise, the distance between every two adjacent white circles can be expressed as 2212221, which can be called the diatonic pattern. This structure describes the pitch intervals between every two consecutive tones in the diatonic scale, and it also describes the temporal intervals between every two consecutive temporal positions in the standard pattern.

The diatonic pattern has two important features. One is that it is asymmetrical (Browne, 1981; Rahn, 1996). Each of the tones bears a unique constellation of relations to every other tone. So even if listeners only hear two or three of them, they can still tell where in this pattern they are, and where the most stable tone or the most stressed tone is. As briefly touched on earlier, the other feature is that the diatonic pattern is cyclic (Iyer, Bilmes, Wright, & Wessel, 1997; Temperley, 2000). This allows a scale or a rhythm to start from any of these seven points and then cycle back to the starting point, which generates seven unique patterns. This theoretically prominent isomorphism between the diatonic scale and the standard pattern suggests a new approach to studying how pitch and time combine perceptually.

In the ongoing debate regarding how information about pitch and time combine, a number of studies have provided evidence for two opposing positions (as summarized in Krumhansl, 2000; see also Prince & Schmuckler, 2014). One is that pitch and time are two separable dimensions, where one dimension does not interact with the perception of the other. In support of this, the perception of melodic similarity has been found to be an additive function of the similarity of the melodic patterns and the similarity of the rhythmic patterns (Monahan & Carterette, 1985). A similar result was found in a study finding that judgments of phrase endings were an additive function of tonal and metrical hierarchies (Palmer & Krumhansl, 1987a&b). Another source of support for the independence position comes from the neuropsychological literature in which patients may loose sensitivity to melodic information while retaining the ability to distinguish between rhythms (e.g., Peretz & Kolinsky, 1993).

The other position is that pitch and time interact perceptually. This means that change in one dimension affects the perception of the other. A corpus study showed that tonally stable pitch classes tend to occur at temporally stable positions (Prince & Schmuckler, 2014). Other studies found a pitch bias where tonal stability affects judgments of temporal positions (Prince, Thompson, & Schmuckler, 2009) and meter perception (Ellis & Jones, 2009). In the opposite direction, studies have found better memory for tones occurring at rhythmically expected points in time (Jones, Moynihan, MacKenzie, & Puente, 2002). Overall, the precise nature of the pitch-time relationship is not yet well understood.

Inspired by the isomorphism between the diatonic scale and the standard pattern, the current experiments took an alternative approach to the question of how pitch and time are processed. The experiments asked whether judgments of how well a rhythm fit a scale could be accounted for by how much the tonally stable tones and the rhythmically stable tones coincided. That is, were the fit judgments higher when the two hierarchies of stability correlated with one another?

II. EXPERIMENT 1

The purpose of this experiment was to measure tonal and rhythmic stability in the seven scales and seven rhythms that are formed by shifting the starting pitch interval or tone duration in Figure 1b. In order to measure the perceived tonal stability of the tones in the scales, Experiment 1a was a probe tone experiment in which each of the probe tones following the scales was judged as to how well into the scale context. In order to measure the perceived rhythmic stability of each of the tones in the rhythms, in Experiment 1b each position of the sequences was accented dynamically and listeners judged how well the accent fit into the rhythm. These judgments were then used in the analysis of the results of Experiment 2 that used all 49 possible combinations of the 7 scales and 7 rhythms in Experiment 1.

A. Method

1) Participants. Forty-five Cornell University students participated in each experiment for course credit or a $5 cash reward. Thirty-one participated in both. In both studies, all but 1 were musically trained. Participants in experiment 1a had an average of 14 years of musical training; 3 had absolute pitch. Participants in experiment 1b had an average of 12.8 years of musical training.

2) Stimulus materials. In both experiments, all sequences consisted of 8 tones, forming 7 intervals. They were created using GarageBand and were sounded in piano timbre. In Experiment 1a, the seven scales were constructed by shifting the starting interval on the diatonic pattern in Figure 1b. For example, Scale1 had the intervals (in semitones) 2212221, Scale2 had the intervals 2122212, and so on. Each scale was constructed in both ascending and descending forms, beginning and ending on C. The range of the ascending sequence was C2 to C3, and the range of the descending sequence was C4 to C5 (details see Krumhansl & Shepard, 1979). The seven scales were followed by a probe tone that was one of the seven tones in the scale, which was played in the range of C3 to C4 between the ranges of the two scale contexts. A baseline trial was also composed for each scale with isochronous rhythm. For Experiment 1b, we constructed seven rhythms by shifting the starting duration on the standard pattern in Figure 1b. For example, Rhythm1 had the durations (in eighth notes) 2212221, Rhythm2 had the durations 2122212, and so on. They were played monotonically on C3. On successive trials, each of the seven temporal positions was dynamically accented; this is called the probe accent. A baseline trial was also composed for each rhythm with monotonic pitch. All rhythms were played twice with a short pause in between.

3) Procedure. Participants were asked to listen to one stimulus at a time and then make their rating. In Experiment 1a, they rated how well the probe tone fits into the scale by moving a slider on a continuous scale from extremely bad fit to extremely good fit. In Experiment 1b, they rated how well the probe accent fits into the rhythm by moving the same slider. In both studies, they first completed four practice trials. The trials were blocked by scale or rhythm. Each block began with the relevant scale or rhythm played in a neutral form, that is, without a probe tone or accent. The neutral form for Experiment 1a was an ascending scale followed by the same scale in descending order, with a short pause in between; the neutral trial for Experiment 1b was a rhythm played twice with a short pause in between. After the neutral trial, they listened to and rated the probes for that particular scale or rhythm, and then moved on to a different scale or rhythm which were presented in a randomized order. Once they rated all probe trials for the seven scales or rhythms, they listened to and rated each of the 7 baseline trials in a randomized order on two scales: how familiar the scale or rhythm is, and how well-formed the scale or rhythm seems (in other words, whether the scale or rhythm forms a good pattern). Both items were also rated by moving a slider on a continuous scale, from extremely unfamiliar or ill-formed to extremely familiar or well-formed. At the end of the study, they filled out a demographics questionnaire. Each study lasted approximately 30 minutes.

III. EXPERIMENT 2

B. Method

4) Participants. Fifty Cornell University students participated in the experiment for course credit. All but 4 were musically trained. Participants had an average of 11.3 years of musical training; 3 had absolute pitch.

5) Stimulus materials. Forty-nine scale/rhythm pairs were constructed by combining the seven scales and the seven rhythms used in Experiment 1. As Table 1 shows, out of all 49 stimuli, the seven on the diagonal in this table are matched pairs, because both the scale and the rhythm start from the same point in the diatonic pattern in Figure 1b. This means that the scale and the rhythm share the same structure. The rest are mismatched pairs, because the scale and the rhythm do not share the same structure. In addition, the same fourteen baseline trials from Experiment 1 were used again as baseline trials.

6) Procedure. Participants were asked to listen to one stimulus at a time. After each listening, they rated how well the rhythm fits the scale by moving a slider on a continuous scale from extremely bad fit to extremely good fit. First, they completed four practice trials, and then rated the 49 pairs in a randomized order. After filling out a demographics questionnaire, they listened to and rated the 49 pairs for a second time, in a different randomized order. After filling out another demographics questionnaire, they listened to and rated familiarity and well-formedness for each of the 14 baseline trials, presented in a randomized order. The study lasted approximately 30 minutes.

IV. RESULTS

Data were processed in the following way. All continuous rating scales were coded from -100 to 100, with -100 being extremely bad fit, unfamiliar, or ill-formed, and 100 being extremely good fit, familiar, or well-formed. For Experiment 1a, probe tone ratings from ascending and descending trials were averaged because they correlated highly with each other. Next, individual ratings were averaged across participants because no large effect of musical training background was found in either Experiment 1a or 1b. This way, one judgment rating was obtained for each probe tone in each scale, and one judgment rating was obtained for each probe accent in each rhythm. Probe tone ratings was then correlated with probe accent ratings, which gives a predicted goodness of fit measure for how participants would judge the combined pitch and time pattern in Experiment 2. Table 2 shows this goodness of fit measure from Experiment 1.

Similarly, in Experiment 2, the two judgment ratings for each scale/rhythm pair were averaged because they correlated highly with each other. Individual ratings were also averaged across participants because no large effect of musical training background was found. Table 3 shows the ratings of how well the rhythm fit the scale for each of the 49 pairs from Experiment 2. To determine how information about pitch and time combine perceptually, we correlated the goodness of fit measure from Experiment 1 with the judgment ratings from Experiment 2.

If pitch and time are separable dimensions, then the exact scale/rhythm combination should not matter for the cross-dimension judgment. In other words, the correlation between probe tone ratings and probe accent ratings should not predict the cross-dimension judgment ratings. Thus, the expected correlation between the goodness of fit measure and the judgment ratings would be zero if the two dimensions are processed separately. On the other hand, if the correlation is not zero and is significant, then it means that pitch and time are not separable dimensions and that they interact in perception.

Results show a positive and significant correlation between the goodness of fit measure and judgment ratings, $r(49) = .65$, $p < .0001$. This suggests that pitch and time interact in the perception of music. They interact in such a way that listeners prefer the higher-rated tones to be played on higher-rated temporal positions.

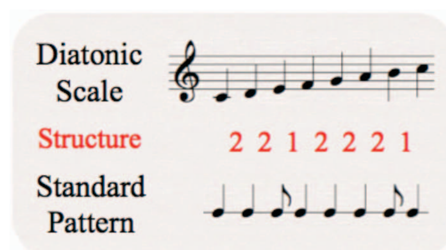
However, listeners in Experiment 2 reported being much more familiar with the major diatonic scale than the others. Consequently, we computed the residuals of listener's judgments after taking out the effect of familiarity. A correlation analysis was conducted to assess the judgments against the goodness of fit measure. Results remain positive and significant, $r(49) = .51$, $p < .001$. This suggests that after taking the familiarity bias into account, listeners still preferred the higher-rated tones to be played on higher-rated temporal positions.

In addition, the judgment ratings were examined against the surface-level structural match between pitch interval and tone duration. As Table 1 shows, all matched pairs were coded as 1 and the rest as 0. Correlation between judgment and surface-level match was not significant, $r(49) = .05$, $p = .75$. The surface-level match was also coded in two other ways. One way was to count the number of times the pitch interval and the time interval matched. The other was how many positions matched before the mismatch. Neither of the codings of surface-level match correlated significantly with the judgments of how well the rhythm matched the scale; for the first coding, $r(49) = .03$, $p = .86$; for the second coding, $r(49) = .17$, $p = .23$. This suggests that the surface-level structural match does not predict judgments of cross-dimension fit. Instead, it is the match between the underlying tonal stability and rhythmic stability of the tones that predicts judgments of fit.

V. CONCLUSION

The current experiments explored the relationship between musical pitch and time by focusing on the isomorphism between pitch interval and tone duration. Specifically, scales and rhythms in the diatonic pattern were used as stimuli. Findings suggest that the surface-level structural match did not predict judgments of the cross-dimension fit. Instead, the correlation between the two probe ratings, measuring tonal stability and rhythmic stability, predicted judgments of how well the rhythm fit the scale. The ratings were higher when the higher-rated tones were played on the higher-rated temporal positions in the probe experiments. This suggests that listeners' cross-dimension judgments were governed by their preference for the best-fitting tones in the diatonic scales to be played on the best-fitting temporal locations in the standard pattern. This finding shows that pitch and time are not two separable dimensions. Instead, they interact when joined together.

a).



b).

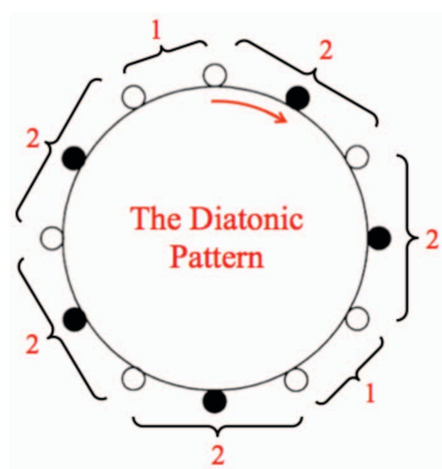


Figure 1. a). Isomorphic structure shared by the pitch intervals of the diatonic scale and the tone durations of the standard pattern. b). Illustration of the asymmetrical and cyclic Diatonic Pattern.

Table 1. Design for Experiment 2 and surface-level coding

	Scale1	Scale2	Scale3	Scale4	Scale5	Scale6	Scale7
Rhythm1	1	0	0	0	0	0	0
Rhythm2	0	1	0	0	0	0	0
Rhythm3	0	0	1	0	0	0	0
Rhythm4	0	0	0	1	0	0	0
Rhythm5	0	0	0	0	1	0	0
Rhythm6	0	0	0	0	0	1	0
Rhythm7	0	0	0	0	0	0	1

Note. Matched pair = 1, mismatched pair = 0

Table 2. Goodness of fit measure constructed by correlating probe tone ratings and probe accent ratings

	Scale1	Scale2	Scale3	Scale4	Scale5	Scale6	Scale7
Rhythm1	.90	.62	.52	.86	.71	.76	.12
Rhythm2	.77	.51	.49	.68	.64	.68	-.05
Rhythm3	.30	.07	.59	.20	.21	.30	.29
Rhythm4	.56	.33	.25	.37	.50	.47	-.17
Rhythm5	.86	.64	.47	.73	.76	.75	.01
Rhythm6	.33	-.16	.22	.45	.06	.08	.14
Rhythm7	.42	.03	.17	.23	.36	.19	-.08

Table 3. Judgment ratings of cross-dimension fit

	Scale1	Scale2	Scale3	Scale4	Scale5	Scale6	Scale7
Rhythm1	27.61	0.39	1.40	4.87	4.61	1.39	-15.95
Rhythm2	41.16	1.81	0.13	9.88	6.55	8.57	-5.83
Rhythm3	13.43	-13.44	-11.24	0.11	-7.96	-9.01	-24.63
Rhythm4	16.69	-0.34	-3.55	12.62	7.00	-5.50	-6.81
Rhythm5	39.27	-0.67	0.17	10.34	6.51	6.30	-9.11
Rhythm6	32.60	-12.54	-5.25	4.71	-9.57	-0.35	-12.99
Rhythm7	10.04	-20.08	-21.19	-7.29	-21.41	-10.13	-24.13

Note. The lowest possible value is -100, meaning extremely bad fit; the highest possible value is 100, meaning extremely good fit.

ACKNOWLEDGMENT

We would like to thank Cognitive Science at Cornell University for the Travel Grant they provided for Olivia Wen in support for her attendance of this conference. We would also like to thank Matthew Shortell for making stimuli and collecting data for Experiment 1.

REFERENCES

Agawu, K. (2006). Structural analysis of cultural analysis? Competing perspectives on the "Standard Pattern" of West African Rhythm. *Journal of the American Musicological Society, 59,* 1–46.

Browne, R. (1981). Tonal implications of the diatonic set. *In Theory Only, 5,* 3–21.

Clough, J., & Douthett, J. (1991). Maximally Even Sets. *Journal of Music Theory, 35,* 93–173.

Ellis, R. J., & Jones, M. R. (2009). The role of accent salience and joint accent structure in meter perception. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 264–280.

Iyer, V., Bilmes, J., Wright, M., & Wessel, D. (1997). A novel representation for rhythmic structure. *In Proceedings of the 23rd International Computer Music Conference,* 97–100.

Jones, M. R., Moynihan, H., Mackenzie, N., & Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological Science, 13,* 313-319.

Krumhansl, C. L. (2000). Rhythm and pitch in music cognition. *Psychological Bulletin, 126,* 159–179.

Krumhansl, C. L., & Shepard, R. N. (1979). Quantification of the hierarchy of tonal functions within a diatonic context. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 579–594.

Monahan, C. B., & Carterette, E. C. (1985). Pitch and duration as determinants of musical space. *Music Perception, 3,* 1–32.

Palmer, C., & Krumhansl, C. L. (1987a). Independent temporal and pitch structures in determination of musical phrases. *Jounal of Experimental Psychology: Human Perception and Performance, 13,* 116–126.

Palmer, C., & Krumhansl, C. L. (1987b). Pitch and temporal contributions to musical phrase perception: Effects of harmony, performance timing, and familiarity. *Perception & Psychophysics, 41,* 505–518.

Peretz, I. & Kolinsky, R. (1993). Boundaries of separability between melody and rhythm in music discrimination: A neuropsychological perspective. *The Quarterly Journal of Experimental Psychology, 46A,* 301–325.

Pressing, J. (1983). Cognitive isomorphisms between pitch and rhythm in world musics: West Africa, the Balkans and Western Tonality. *Studies in Music, 17,* 38–61.

Prince, J. B., & Schmuckler, M. A. (2014). The tonal-metric hierarchy: A corpus analysis. *Music Perception, 31,* 254–270.

Prince, J. B., Thompson, W. F., & Schmuckler, M. A. (2009). Pitch and time, tonality and meter: How do musical dimensions combine? *Journal of Experimental Psychology: Human Perception and Performance, 35,* 1598–1617.

Rahn, J. (1987). Asymmetrical ostinatos in sub-Saharan music: Time, pitch, and cycles reconsidered. *In Theory Only, 9,* no.7: 23–36.

Temperley, D. (2000). Meter and grouping in African music: A review from music theory. *Ethnomusicology, 44,* 65–96.

The effects of timbre on Absolute pitch (AP) judgment

Xiaonuo Li[1], Jiaxuan Yu[2], Diana Deutsch[3]
[1]Department of Musicology, Shanghai Conservatory of Music, China
[2]Department of sociology, Fudan University, China
[3]Department of Psychology, University of California, San Diego, USA

The study reported here capitalized on the high prevalence of absolute pitch (AP) among students in the Shanghai Conservatory of Music (see Deutsch, Li, and Shen, JASA 2013, 134, 3853-3859) to study the effects of timbre on note naming accuracy. The subjects were 71 students at the Shanghai Conservatory of Music; roughly half the subjects were piano majors, the other half being string majors.

The subjects were administered a test which employed the same notes as were employed in earlier studies by Deutsch et al. (Deutsch et al.,2006; Deutsch et al., 2009). They were presented successively with the 36 notes that spanned the three-octave range from C3 (131 Hz) to B5 (988 Hz) and wrote down the name of each note (C, C#, D, etc.) when they heard it. A difference between this and the earlier studies is that we used the 36 notes twice in two timbres: one for a piano timbre and the other for a string timbre. (More specifically, the string timbre consisted of two timbres: the pitches below “G3” were in a viola timbre, and those above “G3”were in violin timbre). The piano and string timbres were tested separately and contained the same numbers of the pitches but indifferent orderings. Table 1 displays the number of subjects in each condition, their majors, and the different orders (piano timbre vs. string timbre) in which the tones were presented.

Table 1. Number of subjects who were piano and string majors and the presentation orders (piano vs. string timbre)

<table>
<tr><th></th><th colspan="3">Group1(P-S)</th><th colspan="3">Group2(S-P)</th><th>Total</th></tr>
<tr><td rowspan="2">Piano Major</td><td colspan="3">13</td><td colspan="3">9</td><td rowspan="2">22</td></tr>
<tr><td>Piano timbre</td><td>Questionnaire</td><td>String timbre</td><td>String timbre</td><td>Questionnaire</td><td>Piano timbre</td></tr>
<tr><td rowspan="2">String Major</td><td colspan="3">27</td><td colspan="3">22</td><td rowspan="2">49</td></tr>
<tr><td>Piano timbre</td><td>Questionnaire</td><td>String timbre</td><td>String timbre</td><td>Questionnaire</td><td>Piano timbre</td></tr>
<tr><td>Total</td><td colspan="3">40</td><td colspan="3">31</td><td>71</td></tr>
</table>

The subjects were piano majors and string majors, and those in each instrument group were divided randomly into two subgroups. One subgroup listened to notes in the piano timbre first, and to notes in the string timbre second; the other subgroup heard the piano and string timbres in reverse order. The subjects also completed a questionnaire concerning their personal information. As in the earlier study (Deutsch, Li, and Shen, 2013), given the low performance level for age-of-onset group ≥ 10, the analyses in this study were carried out taking only those subjects (n = 71) who had begun musical training at age ≤9.

Overall performance levels were very high; the subjects scored 73% correct, not allowing for semitone errors, and 80% correct allowing for semitone errors.

First, we evaluated the basic difference in the average percentage correct between the piano and string timbres. To examine this difference, the subject population was divided into two subgroups, Group 1 heard piano timbre first, and Group 2 heard string timbre first. We compared the percentage correct for piano timbre in Group 1 and string timbre in Group 2. It was found that the percentage correct for piano timbre(the average percentage correct for piano timbre was 76.8% correct not allowing for semitone errors, and 84.15% correct allowing for semitone errors)was higher than for string timbre(the percentage correct for string timbre was 69.25% correct not allowing for semitone errors, and 77.23% correct allowing for semitone errors;) and the difference between piano and string timbre was highly significant[Semitone errors not allowed: $p < .02$; Semitone errors allowed: $p < .01$].

Second, we compared the average percent correct for each timbre separately depending on whether they heard it first or second. The percentage correct for the piano timbre when it was heard first(78.82% correct not allowing for semitone errors, and 86.46% correct allowing for semitone errors)was little higher than when it was heard following the string timbre (74.19% correct not allowing for semitone errors, and 81.18% correct allowing for semitone errors. However, the percentage correct for the string timbre was higher when it was heard following the piano timbre(77.43% correct not allowing for semitone errors, and 85.76% correct allowing for semitone errors) than when it was presented first(58.69% correct not allowing for semitone errors, and 66.22% correct allowing for semitone errors).

The difference in percentage correct between subjects from the different majors was also examined for each timbre. In most cases, the performance of the piano majors (75.69% correct not allowing for semitone errors, and 82.51% correct allowing for semitone errors)was slightly higher than that of the string majors(71.83% correct not allowing for semitone errors, and 79.88% correct allowing for semitone errors); however this difference was nonsignificant ($p>0.1$).

Generally, the performance of all the subgroups was higher when they were judging pitches in the piano timbre than in the string timbre. Further, after listening to the piano timbre, naming the pitches in the string timbre was significantly more accurate. Yet after listening to the string timbre, naming the pitches in the piano timbre was less accurate, though not significantly so. Also, the performance of the piano majors was higher than that of the string majors, though not significantly so. Given this pattern of results, we conclude that the difference in ability to identify pitches in these different timbres found here was not due, in general, to a difference in long term experience with these two types of timbre, but rather to the characteristics of these timbres, and to the context in which they were presented.

Using pattern-classification to uncover the dynamic neural representation of pitch in a tonal context

Narayan Sankaran[1,2], Tom Carlson[2,3], Simon Carlile[1], Lincoln Colling[4], Bill Thompson[2,5]
[1]School of Medical Sciences, University of Sydney, Australia
[2]ARC Centre of Excellence in Cognition and its Disorders, Macquarie University, Australia
[3]Department of Cognitive Sciences, Macquarie University, Australia
[4]Department of Psychology, Centre for Neuroscience in Education, University of Cambridge, UK
[5]Department of Psychology, Macquarie University, Australia

Within a harmonic context, pitch relationships between notes are governed by syntactic rules that instantiate the perception of tonal structure. Krumhansl and colleagues showed that behavioral ratings can be used to derive geometric models of mental representations of musical pitch. Individual pitches are mapped onto a continuum of stability depending on their relationship to a tonal context, and this context determines the perceived similarity between pitches. The current study examined the mental representation of pitch at a neural level using Multivariate Pattern Analysis applied to Magnetoencephalography data, recorded from musically trained subjects. On each trial, one of four harmonic contexts was followed by one of four single probe notes. Two probe notes were diatonic (tonic and dominant) and two were nondiatonic. A machine learning classifier was trained and tested on its ability to predict the stimulus from neural activity at each time-point. We found that pitch-class was decodable from 150 ms post-onset. An analysis of the decodability of individual pairs of pitches revealed that even two physically identical pitches preceded by two distantly related tonal contexts could be decoded by the classifier, suggesting that effects were driven by tonal schema rather than acoustical differences. Characterizing the dissimilarity in activity patterns for each pair of pitches, we uncovered an underlying neural representational structure in which: a) diatonic tones are dissimilar from nondiatonic tones, b) diatonic tones are moderately dissimilar to other diatonic tones, and c) nondiatonic tones are closely related to other nondiatonic tones. Dissimilarities in neural activity were significantly correlated with differences in goodness-of-fit ratings, indicating that psychological measures of pitch perception honor the underlying neural population code. We discuss the implications of our work for understanding the neuroscientific foundations of musical pitch.

A Cross-Modal Comparison of Veridical and Schematic Expectations

Kat Agres[1], Moshe Bar[2] and Marcus Pearce[1]
[1]Department of Electronic Engineering and Computer Science, Queen Mary, University of London, London, UK
[2]Gonda Center for Brain Research, Bar-Ilan University, Ramat Gan, Israel

The topics of expectation and prediction have received considerable attention in music cognition (e.g., Huron, 2006; Pearce, 2005; Temperley, 2007) and visual perception (e.g., Bar, 2007; Summerfield & Egner, 2009; Fiser & Aslin, 2002), but few studies have examined the relative surprise that can result from *veridical expectations*, resulting from exact repetitions of material in the local context, versus *schematic expectations*, acquired through prior exposure. It is also unclear whether the underlying expectation mechanisms are modality-specific or more general. To address these questions, we aimed to test different types of expectation (veridical and schematic) by systematically varying the predictability of artificially constructed stimuli, and compared expectation processing across domains. Every trial consisted of a sequence of tones or visual elements, and participants rated the expectedness of a target (the final event in the sequence). The target either fulfilled or violated expectations set up by the preceding context, which varied in predictability. In both modalities, contexts yielding specific veridical expectations led to more extreme un/expectedness ratings than contexts producing general schematic expectations, and non-repetitive/less predictable stimuli produced the least surprise from an improbable target. In a direct extension of these findings, work in progress presents visual and musical information concurrently to addresses the extent to which veridical and schematic expectations from one modality impact expectations within the other modality. Although different information contributes to the degree of predictability of stimuli across modalities, we conclude that expectations in both music and vision are driven by domain general processing mechanisms that are based on the degree of specificity and predictability of the prior context.

Revisiting music perception research: Issues in measurement, standardization, and modeling

Convenor: Beatriz Ilari[1]
Graziela Bortz[2], Hugo Cogo-Moreira[3], Nayana di Germano[4]
[1]Department of Music Teaching & Learning, University of Southern California, USA; email: ilari@usc.edu
[2]Department of Music, UNESP, São Paulo, Brazil; email: g_bortz@hotmail.com
[3]Department of Psychiatry, Federal University of São Paulo, Brazil; email: hugocogobr@gmail.com
[4]Department of Music, UNESP, São Paulo, Brazil; email: nayanager@hotmail.com

The study of music perception has received considerable attention from the scholarly community. Models of music perception have been devised, and several test batteries created and used, time and again, to investigate how humans make sense of the musical sounds found in their surroundings. In spite of the popularity of many studies and tests of music perception, discussions concerning their underlying epistemological assumptions, measurement properties and standardization are still rare. The aim of this symposium is to revisit music perception research from the perspective of psychometrics. As a contribution for the construction of critical thinking in this area, the participants of this symposium will share findings from three research projects concerned with the measurement of music perception.

Following a brief introduction of the symposium theme and presenters, the first presentation focuses on the construction of a new model for testing absolute pitch. The second presentation describes an experimental study on active and receptive melodic perception in musicians that is being carried out in Brazil. Finally, the third presentation brings forward the issue of construct validity in music perception research, following an analysis of 163 studies. The discussant will make some comments and then there will be time for discussion with audience members.

Absolute Pitch: In Search of a Testable Model

Nayana Di Giuseppe Germano,*[1] Hugo Cogo-Moreira,#[2] Graziela Bortz*[3]

*Department of Music., UNESP, Sao Paulo Brazil

#Department of Psychiatry., Federal University of Sao Paulo, Sao Paulo Brazil

[1]nayanager@hotmail.com, [2]hugocogobr@gmail.com, [3]g_bortz@hotmail.com

ABSTRACT

After an extensive bibliographical review regarding the cognitive trait known as absolute pitch (AP), we observed that its main characteristic lies in the ability to recognize and identify tones using verbal labels without any kind of external reference. However, its several definitions also include various non-consensual criteria. Therefore, it is important to find a proper approach to cope with the indicators that classify AP, establishing cut-off points with accuracy rates to distinguish those who have AP from those who do not have it. The primary purpose is to suggest a set of indicators to assess what AP is, so that the resulting theoretical model could be tested in a future research phase.

The reproducible model for AP (which consists in the proposed indicators) can be established by using standardization of criteria and construction of validity evaluation, as conducted in the medical area through guidelines such as *Diagnostic and Statistical Manual* of *Mental Disorders* or *Composite International Diagnostic Interview*. The test of the resulting theoretical model shall use procedures from a field of statistics known as structural equation modelling, focused in testing theoretical models. The existence of diverse definitions for the same latent phenomenon using different criteria inevitably leads to distinct and non-directly comparable models. For the study of any latent psychological trait, it is essential to identify a set of observable indicators (e.g., criteria, expressions). Such criteria must meet content validity based on evidence from empirical observation and theoretical foundation. This is an important step for AP researchers, since an adequate testable model for this cognitive trait must be defined and consequently evaluated regarding their fit with reality.

I. INTRODUCTION

The Absolute Pitch phenomenon (AP) has been widely researched and discussed during the last century. The earliest scientific description of AP appeared in a volume on psycho-acoustics by Stumpf (1883). Since then, various aspects of AP have been investigated by a number of researchers (e.g., Abraham, 1901; Wellek, 1938; for reviews, see Takeuchi & Hulse, 1993; Ward, 1999 *apud* Miyazaki & Ogawa, 2006).

From a general perspective, the AP phenomenon is usually defined (with minor variations among different authors) as a rare ability that refers to a long-term internal representation for the pitches of tones. It is typically manifested behaviorally by the ability to identify, by the name of the musical note, the pitch of any sound without reference to another sound or by producing a given musical tone on demand without external reference (Baggaley, 1974; Zatorre et. al., 1998; Ward, 1999; Parncutt & Levitin, 2001; Deutsch, 2002).

After an extensive bibliographical review of AP definitions conducted by several researchers (Germano, 2015), it has been observed that the main characteristic of this cognitive ability is the capacity that AP possessors have of tone identification using verbal labels (what Levitin, 1994, named as *Pitch Labeling*) without reference (e.g., without a diapason). On the other hand, by trying to delineate a more accurate description, some authors also include more specific and non-consensual criteria to the definition of AP. To illustrate these criteria, we can quote the time demanded to identify a tone and the degree of precision in tone identification (directly questioning the common sense which establishes that if a subject has problems with tone identification due to certain musical parameters, like register and timbre, he should be disqualified as an AP possessor).

The existence of several definitions for the same cognitive phenomenon, each based on different non-consensual criteria, inevitably leads to significant variations among AP researches and, more importantly, in their conclusions. For example, there is a general, although not universal, agreement that, to be considered an AP possessor, an individual must have the ability to recognize pitches immediately and involuntarily (e.g., Takeuchi & Hulse, 1993), however, in one of our previous studies (Germano et al., 2011), some self-reported AP possessors described the need of some time to identify them. There are also disparities regarding the degree of precision of AP possessors, as some authors consider half-tone errors as partial successes (e.g., Athos et al., 2007), while others consider halt-tone errors as total successes considering subjects older than 45 years old (e.g., Baharloo et al., 1998) and others consider total successes for any subject (e.g., Brady, 1970; Schulze et al., 2009; Vanzella & Schellenberg, 2010).

The issue with non-consensual criteria is that they might lead to different AP definitions, which in turn lead to different theoretical models for the cognitive ability known as AP. As a result, one should question if these researches are in fact evaluating the same cognitive ability. In fact, one of the main problems regarding AP research is the difficulty that comes from the adoption of different and non-directly comparable models, making it nearly impossible to cross-relate the experimental results of these researches.

Consequently, the primary purpose of this paper is to propose a set of indicators to assess what AP is, so that the resulting theoretical model could be tested in a future research phase.

II. ABSOLUTE PITCH POSSESSOR

In light of the considerations above, a few questions can be listed, such as: what are the necessary criteria to define AP? Considering a set of defined criteria, are they able to explain, adequately and satisfactorily, the underlining latent psychological trait?

One of the difficulties regarding AP meaning is the amplitude of its definition. If one adopts a very restrictive definition (e.g., AP possessors must identify precisely every note on every instrument), there would be only a few subjects successfully classified as AP possessors. On the other hand, a broader definition (e.g., if AP possessors can take as much time as needed to identify a note or if they are freely allowed prone to semitone errors) can make it very difficult, or even impossible, to distinguish AP possessors from non-possessors, since so many answers are considered indicatives of the cognitive ability.

To ponder on AP definition it is crucial to a scientific approach of the phenomenon. Without a proper approach, it is not even possible to claim that the dichotomous theoretical model commonly adopted to represent AP is in fact adequate. Observing current researches, it is possible to enumerate criteria used to address AP ability, which do not fit into a dichotomous model, although this model is usually presented on the given definition of the cognitive trait. Instead, these authors should have evaluated alternative models for representing AP (e.g., if the phenomenon is better represented by a continuous model or even a hybrid model).

The previous paragraph points out to a need of a deeper thinking regarding AP definition. In fact, many papers define AP generically, while others only quote definitions presented by reference authors, usually without any consideration of the correspondence of these definitions to the AP perspective discussed on these papers.

A better comprehension of the AP phenomenon can begin to take form by observing the consensual criteria identified on AP definitions of different authors. First, AP possessors are capable of identifying pitches immediately, or almost immediately. Second, AP possessors are capable of associating pitches to verbal labels, an association stored on subject´s long-term memory, that is, a specific label is learned by the subject (i.e., usually a name of one of the musical tones) in association to a specific pitch class. Although these associations are considered standardized by many, it is well known that some subjects (e.g., musicians who play transposable instruments or that studied on an out of tune piano) usually show different patterns of associations than most musicians.

Nonetheless, a vast number of researches highlight that AP possessors are rarely infallible, that is, most of them show a few (or many) limitations, such as:

1) Margin of errors of semitone on pitch evaluation (i.e., half-tone errors are recurrent);

2) Difficulty in (or even incapacity of) singing a note without external reference;

3) Limitations to certain registers;

4) Limitations to certain timbres.

If many authors discuss these criteria, the question remains as to how should the scientific community cope with these variables. For example, what is the percentage of correct answers a subject must achieve to be considered an AP possessor? An AP possessor should be capable of identifying pitches on which timbres? If a subject only recognizes pitches on the medium register, should he be considered an AP possessor? If the subject cannot sing a demanded pitch without external reference, can he be considered an AP possessor?

Considering the above information, maybe one could contemplate the cognitive trait named AP heterogeneous according to the subject´s abilities. This leads to a hypothesis (already raised by previous researchers, as Bachem in 1937) that there could be different types of AP possessors.

III. MODEL EVALUATION

To study any latent psychological phenomenon (i.e., one that cannot be measured directly, such as AP), the identification of a set of structured, consistent, observable criteria based on evidence is essential. From a scientific perspective, it is of prior importance that these elaborated criteria can be tested in search for evidences supporting the theoretical model elected. It is from this experimental testing that the model can be rejected or not, allowing its enhancement.

Most AP researches adopt experimental testing as their core methodology, in order to measure different patterns of subject response in relation to parameter variation (such as register, timbre, the time necessary for pitch identification or the proportion of correct/wrong answers). However, as highlighted in the previous section, the lack of consideration regarding correspondences between theoretical model and empirical data makes the process of knowledge acquisition extremely difficult.

If how well the proposed set of criteria for the definition of the AP phenomenon fits to the reality of the AP possessor and its abilities is unknown, then solving this matter should be the first phase in experimental research. In fact, the borderline that separates AP possessors from non-possessors has not been consensually defined, taking into account that both exhibit limitations on pitch identification. Considering that many researches adopt, as a starting point of their methodology, the segregation of AP possessors from non-possessors, this problem becomes of great importance.

Consequently, the scientific community should find a way to cope with the basic indicators used to classify AP, defining them properly and establishing cut-off points with accuracy rates, in order to separate those who have AP from those who do not have it. This demands the accomplishment of experiments specifically designed to test basic hypothesis on AP ability, based both on evidence from previous empirical observations and on logical coherence provided by theoretical foundations.

Only when this basic step is achieved will researches dedicated to more specific aspects of AP ability have a solid ground to elaborate on. As pointed before, without the standardization of basic criteria inside the community dedicated to the AP studies, it is not possible to know if different researches are in fact evaluating the same cognitive ability, and even less to compare information resulting from these experiments.

A possible solution to this problem is the creation of a reproducible model for AP categorization using standardization of criteria and the construct validity evaluation as conducted in the medical area, throughout guidelines stated in sources such as *Diagnostic and Statistical Manual* of *Mental Disorders* or *Composite International Diagnostic Interview*. The resulting theoretical model could then be tested by using procedures from a field of statistics known as structural equation modelling, focused specifically on testing theoretical models.

IV. ABSOLUTE PITCH INDICATORS

What are the criteria necessary to define AP? Considering a designed set of criteria, are they capable of adequately explaining the latent psychological trait?

From a scientific perspective, it is important that the designed criteria can be tested in order to provide evidences supporting or falsifying the core theoretical model adopted, allowing its enhancement.

Based on researched bibliography, more specifically those discussed in Germano (2015), we propose a preliminary hypothetical model (construct) for AP:

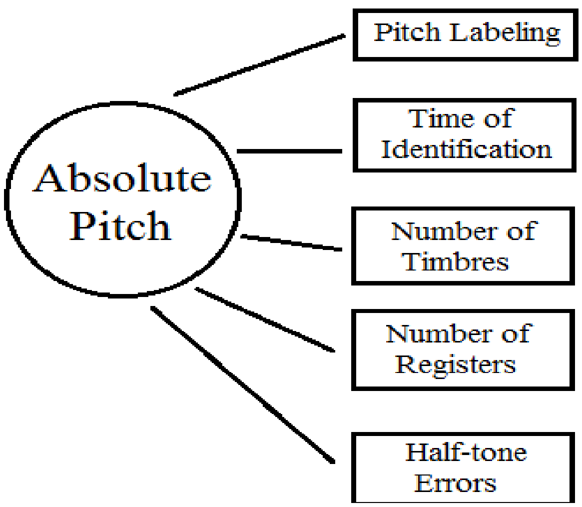


Figure 1. Preliminary hypothetical model for AP.

This model is a starting point for future researches. Therefore, it is open to modifications and corrections due to results obtained within the preliminary test that should evaluate this hypothetical proposition.

From a theoretical perspective, the proposed model considers that AP is a dichotomous phenomenon (to possess or not to possess) and embodies the following indicators:

1) Pitch Labeling: ability to keep long-term stable representations of specific pitches on memory and to access them when needed, associating them to learned verbal labels (Levitin, 1994).

2) Time of identification: according to AP bibliography, the identification of stimuli in AP possessors is immediate. Our question is: what is immediate? How many seconds does the subject need to identify a tone?

3 and 4) Number of timbres and registers: these items measure the capacity of the subject to identify pitches in two parameters, describing how many timbres and registers a subject is capable of recognizing (from identifying in just one timbre and one register to all timbres and their respective registers).

5) Half-tone errors: how many half-tone errors does an AP possessor make? The lower degree of accuracy would be 100% half-tone errors and the higher degree would be 0% half-tone errors. However, it should be questioned if a large quantity of half-tone errors is actually caused by the subject´s mistakes, or if it is due to a different memorization reference. As pointed out before, many studies highlight that AP possessors acquire pitch memory according to the reference presented to them on the first years of study. For example, if a child possessing AP grows studying in an out of tune piano (tuned half-tone bellow the A=440 Hz standard), this child will certainly make consistent half-tone errors in tests, although the variation in errors would possibly be near zero (Levitin & Rogers, 2005). Due to this fact, Ward considers that the best AP possessor is not the one who makes fewer mistakes, but the one who exhibits fewer variations in answers (WARD, 1999).

Although we have adopted a dichotomous perspective, it is important to emphasize that the proposed indicators can collectively provide a mapping of the cognitive abilities of AP possessors in all its diverse capacities and limitations, therefore providing a model which is believed to be better capable to fit the reality of AP possessors' abilities.

Consequently, this hypothetical model can provide a solid perspective to AP phenomenon. Through test, it can supply solid evidences to indicate if AP can be described as a continuous ability (with a continuous line from minimal to maximal pitch identification), or if it is better described as a hybrid ability (with the possible verification of different AP types, as already proposed by Bachem).

V. CONCLUSION

The primary purpose of this paper was to propose a set of indicators to assess what AP is, so that the resulting theoretical model could be tested in a future research phase. The elaboration of this model started by observing the consensual criteria present on AP definitions of different authors.

Based on the information provided, the scientific question of what in fact is the AP ability and how it has been measured remains open on researches up to date. We highlighted the main non-consensual AP characteristics, which show that AP possessors commonly exhibit some kind of limitation (be it regarding timbre, register or half-tone errors). We also questioned the common (undiscussed) view that AP is a dichotomous cognitive trait.

Based on researched bibliography, more specifically those discussed in Germano (2015), we proposed a preliminary hypothetical model (construct) for AP. First, it will be evaluated, leading to modifications and corrections according to results obtained within the preliminary test aimed at testing this hypothetical proposition. As mentioned before, the elaboration of a well-formulated (testable) model for AP categorization, formed by a coherent set of criteria, is an essential step for AP researchers. It is only by its test that the scientific community will be able to evaluate the capacity of the model to fit reality, i.e., the mapping of the cognitive abilities of AP possessors in all its diverse capacities and also limitations.

REFERENCES

Athos, E. A., Levinson, B., Kistler, A., Zemansky, J., Bostrom, A., Freimer, N., & Gitschier, J. (2007). Dichotomy and perceptual distortions in absolute pitch ability. *Proceedings of the National Academy of Sciences*, 104 (37), 14795–14800.

Abraham, O. (1901). Das absolute Tonbewusstsein. Psychologisch-musikalische Studie. *Sammelbände der internationalen Musikgesellschaft*, 3, 1-86.

Bachem, A. (1937). Various types of absolute pitch. *Journal of the Acoustical Society of America*, 9, 146-151.

Baggaley, J. (1974). Measurement of absolute pitch. *Psychology of Music*, 2 (2), 11-17.

Baharloo, S., Johnston, P. A., Service, S. K., Gitschier, J., & Freimer, N.B. (1998). Absolute Pitch: An Approach for Identification of Genetic and Nongenetic Components. *The American Journal of Human Genetics*, 62 (2), 224–231.

Brady, P. T. (1970). Fixed-scale mechanism of absolute pitch. *Journal of the Acoustical Society of America*, 48, 883-887.

Deutsch, D. (2002). The Puzzle of Absolute Pitch. *Current Directions in Psychological Science*, 11 (6), 200-204.

Germano, N. G. (2011). Categorização de Ouvido Absoluto em Estudantes de Música de Nível Universitário nas cidades de São Paulo e Brasília. *Bachelor's Dissertation – Instituto de Artes da UNESP*. São Paulo, november 2011.

Germano, N. G. (2015). Em busca de uma definição para o Fenômeno do Ouvido Absoluto. *Master's Dissertation - Instituto de Artes da UNESP*. São Paulo, june 2015.

Levitin, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of learned memories. *Perception and Psychophysics*, 56 (4), 414–423.

Levitin, D., & Rogers, S. (2005). Absolute pitch: perception, coding, and controversies. *Trends in Cognitive Sciences*, 9 (1), 26-33.

Miyazaki, K. (1988). Musical pitch identification by absolute pitch possessors. *Perception e Psychophysics*, 44 (6), 501–512.

Miyazaki, K., & Ogawa, Y. (2006). Learning Absolute Pitch by Children. *Music Perception*, 24 (1), 63-78.

Parncutt, R., & Levitin, D. J. (2001) Absolute pitch. In S. Sadie (Ed.), *New Grove Dictionary of Music and Musicians* (pp. 37-39). New York: St. Martins Press.

Profita, J., & Bidder, T. G. (1988). Perfect Pitch. *American Journal of Medical Genetics*, 29 (4), 763–771.

Stumpf, C. (1883). *Tonpsychologie (v. 1)*. Leipzig: Verlag von S. Hirzel.

Schulze, K., Gaab, N., & Schlaug, G. (2009). Perceiving pitch absolutely: Comparing absolute and relative pitch possessors in a pitch memory task. *BMC Neuroscience*, 10 (1), 10-106.

Takeuchi, A. H., & Hulse, S. H. (1993). Absolute Pitch. *Psychological Bulletin*, 113 (2), 345–361.

Vanzella, P. & Schellenberg, E. G. (2010). Absolute Pitch: Effects of Timbre on Note-Naming Ability. *PloS one*, 5 (11), e15449.

Ward, W. D. (1999). Absolute Pitch. In D. Deutsch (Ed.), *The Psychology of Music (2nd Edition)* (pp. 265-298). San Diego: Academic Press.

Wellek, A. (1938). Das absolute Gehör und seine Typen [Absolute pitch and its types]. *Zeitschrift für angewandte Psychologie und Charakterkunde-Beihefte*, 83, 1–368.

Zatorre, R. J., Perry, D. W., Beckett, C. A., Westbury, C. F., & Evans, A. C. (1998). Functional anatomy of musical processing in listeners with absolute pitch and relative pitch. *Proceedings of the National Academy of Sciences of the United States of America*, 95 (6), 3172-3177.

Active and Receptive Melodic Perception: Evaluation and Validation of Criteria

Graziela Bortz,*1 Hugo Cogo-Moreira,#2

* *Department of Music, UNESP Sao Paulo, Brazil*
# *Department of Psychiatry, Federal University of Sao Paulo, Brazil*
[1]g_bortz@hotmail.com, [2]hugocogobr@gmail.com

ABSTRACT

Background: Many music theorists and researchers believe that the ability to decode an isolated interval does not necessarily result in differentiating this same interval in real music. Yet there are no construct validity studies investigating the psychometric features of the used tests, neither randomized experimental trails to evaluate the effectiveness of interval recognition training on the perception of melody as Gestalt. Aims: a) evaluate the model underlying items composed to assess melodic perception ability; b) estimate items difficulties and discrimination for receptive and active melodic perception of musicians, and c) estimate the degree of agreement between three judges regarding the criteria for the active melodic evaluation. Method: Sample is constituted by 150 students of an undergraduate school of music for the receptive evaluation. For the productive evaluation trait, out of 150, a random subsample of 50 students is selected and is being evaluated by three judges, according to four items under a five-point-Likert structure. The agreement regarding specialists' ratings on the active melodic evaluation domain will be assessed via weighted Kappa coefficient. To evaluate the construct validity, confirmatory factor analysis (CFA) will be used, estimating a) parameters of difficulty and discrimination for each item within each domain (dictation and sight-singing as a melodic perception trait of musicians), and b) the amount of latent trait for each ability. Results: Research is in progress. Although data collection is finished, evaluation by the judges is being done. Discussion: Music students have being graded in ear-training tests with no tool (test/scale/battery) showing evidences about construct validity that allows us to measure how good the latent trait of perception ability behaves. It is important to provide psychometrical evidences regarding the tools used to assess musical skills.

I. INTRODUCTION

A. Intervals and melodic cognition

Many music theorists and researchers (Karpinsky, 2000; Rodgers, 1996; Covington & Lord, 1994; Edlung, 1963, 1974) believe that the ability to decode an isolated interval does not necessarily result in transferring this same perception into real music. Even in post-tonal contexts, Edlung (1963) always propose sight-reading drills in "chunks". At least two consecutive intervals are presented in varied directions and orders, or the dividing interval (the inside interval as opposed to outer interval) is varied. Covington & Lord (1994, 162) say that the fact a music student is able, for instance, to perceive an isolated tritone does not mean that he or she will automatically be able to interpret that same interval in a music context in which it appears as being formed by the fourth and seventh degrees of a tonal scale, hence being part of a dominant seventh chord, or by the second and sixth degrees of a minor tonal scale, carrying out the function of an intermediate harmony (II chord as a substitute subdominant harmony). "Isolating elements from their natural context emphasizes the separateness of the elements rather than their integration." They add:

> Drilling sonorities separately from their context may cognitively link them more to other isolated sonorities rather than to the roles each sonority might play in a phrase. In fact, research has shown in other domains that such training can actually develop barriers between the schema types instead of developing an awareness of their interconnectedness (Covington & Lord, 1994, 162).

Karpinsky (2000) mentions how vocal sight-reading in tonal music involves a lot more than decoding orders of tones, and Rodgers (1996, 149) discusses the advantages of the "scale-degree function" as being "especially effective for teaching such tonal bearings—a method that stresses the tendency-tone and resolution patterns that define the pitch centricity found within the major/minor tonal system."

The strong tonal centricity appears in tonal hierarchies' cognition studies of Krumhansl et al (1990). They experiment single probe tones and chords related to a musical context, and their findings correspond to musical theory expectations: in the case of tones, in major keys, in fitting into the context, the tonic note receives the highest ratings; in minor keys, the tonic is also rated as the best fitted, followed by the third degree of the scale, hence carrying ambiguities. In the case of chords, the tonic triad receives the highest ratings, followed by the subdominant and dominant chords. Krumhansl & Toiviainen (2009, 107) say that "the sense of key develops and strengthens as certain cues appear, then may weaken or shift to a new key as subsequent events are sounded."

Bigand & Poulin-Charronnat (2009, 63-67) discuss two types of tonal models: Lerdhal's model (1988), and Bharucha's model (1987). They add that, as formulated by Lerdahl & Jackendoff's (1983), "the tonal hierarchy provides the stability conditions that need to be integrated into rhythmic parameters (notably) in order to define the structural importance of each tone of the piece" (p. 69).

Because tonal centricity is so pervasive, it seems that the musician's brain will access such information (scale-degree hierarchies) while doing a solfege much faster than thinking it in little pieces. Yet there are no construct validity studies investigating the psychometric features of the used tests, neither randomized experimental trails to evaluate the effectiveness of interval recognition training on the perception of melody as Gestalt.

The first part of this work addresses interval performance of students in two moments: 1) while sight-reading (active interval performance); 2) in dictation (receptive interval performance). On the other hand, performances on both active and receptive perception of intervals will be compared to active melodic performance (sight-reading of tonal melodies), as well as to receptive melodic performance (transcription of dictated tonal melodies).

II. EVALUATION AND ASSESSMENT: CREATING A BATTERY AND ITS UNDERLYING MODEL

The second part of the present research consists on designing a model (construct) for assessment of the active tonal melodic perception (solfege), and giving criteria for evaluation of such ability. The questions made are: a) How do we grade music students on tonal sight-singing? b) Which are the items/indicators (criteria) that should compose a test?

The Item Response Theory (IRT), especially that described by Lord (1952), and Rasch (1960) is used here. “Performance is the effect, and latent models are the cause.” … “Subjects with greater ability will have the most probability to respond correctly to the item, and vice-versa” (Pasquali, 2003, 82-83).

A theoretical model has been conceived over several years of experience on ear-training classes and evaluation of students at college level, as well as on existing literature. The hypothetical model created here (Figure 1) refers to melodic perception of humans in general.

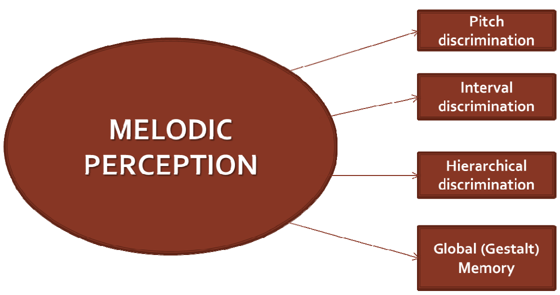


Figure 1. Theoretical Model for melodic perception.

The model above takes into consideration both musicians and non-musicians abilities, because melodic perception is a trait innate to everybody, like intelligence and memory. The item “Pitch discrimination” embeds concepts of “pitch memory” and “pitch labeling” as described by Levitin (1994) for non-musicians and musicians, respectively. Hierarchical discrimination” embeds both tonal and modal systems, and post-tonal different pitch organizations using 12-tone Western music (pitch-class sets).

The model showed in Figure 2 interprets how Western music students traditionally have been evaluated in ear-training classes:

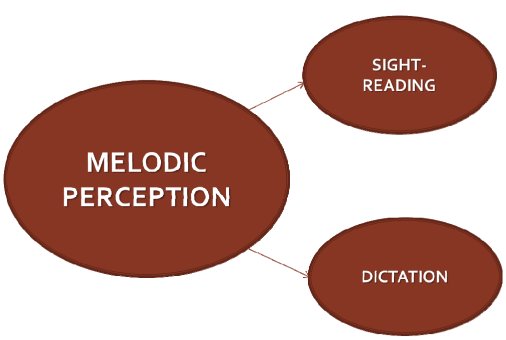
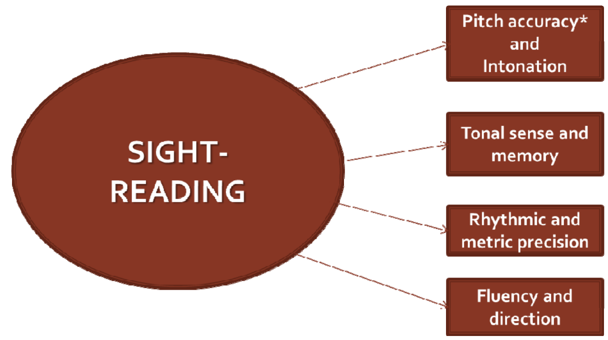


Figure 2. Evaluation of abilities of music students on melodic perception of Western music

The model of Figure 3 refers to the items conceived here to address the Sight-Reading Trait underlying melodic perception of classical music students:



* Pitch accuracy includes the ability of labelling (be it with syllables, numbers or else)

Figure 3. Items that compose the test for evaluating the Sight-Reading Trait of classical music students.

Hence, the Latent Model for such trait can be viewed at the following table, where the lines describe the items used for the test, and the columns refer to the amount of ability on each item, according to a five-point-Likert scale:

Table 1. Items and amount of Latent Trait of the test.

Tonal melodic solfege	A	B	C	D	E
Intonation and pitch accuracy					
Tonal sense and memory					
Rhythmic precision, regularity of pulse and subdivisions					
Fluency and music direction					

III. AIMS AND METHOD

The aims are:

a) Evaluate the model underlying items composed to assess melodic perception ability;
b) Estimate items difficulties and discrimination for receptive and active melodic perception of musicians; and
c) Validate the degree of agreement of criteria for the productive evaluation.

Sample is constituted by 150 students of an undergraduate school of music for the receptive and active evaluation. For the receptive evaluation, 19 intervals and 10 tonal melodies are recorded and dictated to subjects. For the intervals, their answers were classified as dichotomous (correct/incorrect). For each tonal melody, from students’ transcription, the raw score of correct notes note was used as observed indicator to compose the latent trait (10 indicators, then). Rhythmic answers are not considered for this work, although subjects write them on music paper.

For active melodic perception trait, the same 150 students were evaluated by one of the three judges, according to 4 items under a five-point-Likert structure considering 10 passages. The judges were blinded to the students' performance, receiving only a recording with a random ID number.

Students are recorded while sight-reading 19 intervals and 10 tonal melodies (different from the receptive test). The intervals are considered either correct, incorrect, or out of tune by the judges.

Out of 150, 50 students' recorders, at random, were given to all the judges in order to evaluate the agreement regarding specialists' ratings on the active evaluation domain. Such ratings will be assessed via weighted Kappa coefficient. To evaluate the construct (factorial) validity, confirmatory factor analysis (CFA) will be used. The models will be evaluated regarding their fit using the following fit indices and their respective cut-off: chi-square (χ^2), Confirmatory Fit Indices (CFI), the Tucker-Lewis index (TLI), and root mean square error approximation (RMSEA).

For the evaluation of melodies, in order to minimize subjectivity, a 'translation' of the Likert scale into musical terms for each criterion was provided to the judges, as follows:

- Intonation and pitch accuracy:

One should not consider the kind of precision expected as if singers were being judged by a specialist jury. Instead, it should be considered that students from diverse backgrounds in music, such as instrumentalists, conductors, composers, or music educators are being evaluated. In other words, the degree of vocal technique expected is not that of a singer, but one should be able to recognize the pitches of the melody sung.

For the 5-point-Lickert scale, one should consider:

A. It is not possible to recognize the melody sung at all.
B. It is possible to recognize the melody, but the intonation drops (or raises) at least a half step.
C. It is possible to recognize the melody, but the intonation drops (or raises) up to a half step.
D. It is possible to recognize the melody, but the intonation drops (or raises) less than a half step.
E. It is possible to recognize the melody fully.

- Tonal sense and memory

Although related to item 1 ("intonation and pitch accuracy"), it was conceived to approach a different kind of memory, for instance, when an individual gets lost in terms of pitch during the solfege of a tonal or modal melody, or makes errors of decoding pitches, but is still capable of recovering through tonal memory or pitches just sung.

For the 5-point-Lickert scale, one should consider:

A. It is not possible to recognize the tonality(ies) sung at all.
B. It is possible to recognize the tonic (and eventually the dominant) of the tonality(ies) only.
C. It is possible to recognize the tonality(ies), but occasional errors of tones occur.
D. It is possible to recognize the tonality(ies) and diatonic tones, but chromaticism (chromatic inflexions of diatonic tones) are not realized.
E. It is possible to recognize the tonality(ies) fully, even in case where subtle instabilities of intonation occur during the solfege.

- Rhythmic precision, regularity of pulse and subdivisions

Understood here as the capacity of participants on decoding rhythms, meters, and on maintaining pulse and subdivisions regularity during all the exercise. It is not expected, naturally, that the metronomic speed be maintained during the whole solfege, but, even though changes of speed at the pulse level may occur, they are supported by changes at the subdivision levels. Actually, such a thing is quite desirable, musically, since it denotes control and experience by the participant (on doing *ritardandos* at the end of a phrase, for instance).

For the 5-point-Lickert scale, one should consider:

A. It is not possible, absolutely, to discern both meter(s) and regularity of pulses and subdivisions.
B. It is possible to discern meter(s), but subdivisions are inconsistent (for example, a participant executes a ternary subdivision instead of a given binary one, and vice-versa).
C. It is possible to discern meter(s) and its(their) hierarchical subdivisions, but many inconsistences (errors) occur in the course of the solfege.
D. It is possible to discern meter(s) and its(their) hierarchical subdivisions, but occasional inconsistences occur in the course of the solfege.
E. It is perfectly possible to discern meter(s) and its(their) hierarchical subdivisions, and, when there are flotations of speed, they are supported proportionally by the hierarchical levels of division and subdivisions of the beat (rubato, rit.).

- Fluency and music direction

It refers not only to the capacity of the participant on moving forward regardless of mistakes, but also on doing so (with or without mistakes) in a way that the musical discourse is clearly expressed. The expression *musical discourse* should be understood here as the capacity of the student to prevent inadequate accentuations, being them related to meter, rhythm, or harmony, that prevent the fully expression of musical prosody (taken metaphorically here in relation to musical accents, since it doesn't refer to words, but exclusively to music).

For the 5-point-Lickert scale, one should consider:

A. It is not possible to follow the musical discourse.
B. It is possible to follow the musical discourse, but many interruptions occur.
C. It is possible to follow the musical discourse, but tiny non-intentional delays occur (a brief non-desired fermata, for example, while the student 'takes time' in order to decode pitches and rhythms).
D. It is possible to follow the musical discourse, but extra-accents occur that prevent its total fluency.
E. It is possible to follow the musical discourse fully.

IV. CONCLUSION

Research is in progress. Data collection is finished, as well as evaluation made by the judges. Data is currently being inserted in databases' system, and will be analyzed in the next few months.

Music students have being graded in ear-training tests with no tool (test/scale/battery) showing evidences about construct validity that allows us to measure how good the latent trait of perception ability behaves. It is important to build assessment instruments to measure such ability well. We hope that such a procedure will allow verifying the amount of latent trait of this ability (sight-singing and dictation as a melodic perception trait of musicians) presented in students of music.

ACKNOWLEDGMENT

We thank *Fapesp Funding Agency* at Sao Paulo, Brazil, for supporting this research.

REFERENCES

Bharucha, J. J. (1987). Anchoring effects in music: The resolution of dissonance. *Cognitive Psychology, 16*, 4, 485-518.

Bigand, E. & Poulin-Charronnat, B. (2009). Tonal cognition. In S. Hallam, I. Cross & M. Thaut (Eds.), *The Oxford Handbook of Music Psychology* (pp. 59-71). Oxford: Oxford University Press.

Covington, K. & Lord. C. H. (1994). Epistemology and procedure in aural training: in search of a unification of music cognitive. *Music Theory Spectrum 16,* 2, 159-70.

Edlung, L. *Modus novus*. (1963). Stockholm: Nordiska musikförlaget.

Edlung, L. (1974). *Modus vetus: sight singing and ear-training in major/minor tonality*. Stockholm: Nordiska musikförlaget.

Karpinsky, G. S. (2000). *Aural skills acquisition*. Oxford: Oxford University Press.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. Oxford: Oxford University Press.

Krumhansl, C. L. (2006) Ritmo e altura na cognição musical. In: B. S. Ilari (Ed.) *Em Busca da Mente Musical: Ensaios sobre os processos cognitivos em música–da percepção à produção* (pp. 45-109). Curitiba, PR: Editora UFPR.

Krumhansl, C. L. & Toiviainen, P. (2009). Tonal cognition. In I. Peretz & R. J. Zatorre (Eds.), *The Cognitive Neuroscience of Music* (pp. 95-108). Oxford: Oxford University Press.

Lerdahl, F. (1988). Tonal pitch space. *Music Perception 5*, 3, 315-49.

Lerdahl, F. & Jackendoff, R. (1983). *A Generative theory of tonal music*. Cambridge, MA: MIT Press.

Levitin, D. (1994). Absolute memory for musical pitch: evidence from the production of learned melodies. *Perception & Psychophysics 56*, 4, 414-423.

Lord, F. M. (1952). A theory of tests scores*. Psychometric Mpnograph 7.* Iowa, IA: Psychometric Society.

Pasquali, L. (2003). *Psicomentria: teoria dos testes na psicologia e educação*. São Paulo, SP: Editora Vozes.

Rash, G. (1960). *Probabilistic model for some intelligence and attainment tests.* Copenhagen: Danish Institute for Educational Research, and St. Paul, MN: Assessment Systems Corporation.

Rodgers, M. (1996). The Jersild approach: A sightsinging method from Denmark. *College Music Symposium 36*, 149-61. Retrieved April 11, 2016, from http://www.jstor.org/stable/40374290

Tillman, B., Bharucha, J. J.; Bigand, E. (2000). Implicit learning of tonality: a self-organizing approach. *Psychological Review 107*, 4, 885-913.

Urbina, S. (2007). *Fundamentos da testagem psicológica*. São Paulo, SP: Artmed.

How are we measuring music perception and cognition? Lack of internal construct validity, overuse of row scores, and misuses of Cronbach's Alpha

Hugo Cogo-Moreira[1]
[1]Department of Psychiatry, Federal University of São Paulo, São Paulo, Brazil

Background
Different tasks, batteries, tests, and paradigms are largely used in experimental branch of researches on music perception and music cognition, assessing musical skills (or inabilities as amusia). They aim to provide observed indicators underlying given ability/inability.

Aims
To conduct a narrative review searching evidences for construct validity of the commonly used tasks, batteries, and tests normally used as correlational indicator for biological indicators as brain region's thickness, volume, functionally or other phenotypical behavior/skills (e.g., attention).

Methods
In the Pubmed database, without period restriction, the following term were searched: "music perception" and "assessment".

Results
Out of 163 studies, less than 1% described statistical analysis procedures proper to evaluate the batteries/tests' construct validity. Even commonly used tools as Gordon's Musical Aptitude Profile, Seashore Measures of Musical Talents, and Montreal Battery of Evaluation of Amusia (MBEA) do not present evidence regarding construct validity, which is assessed by factor analysis techniques (e.g., confirmatory factor analysis [under dichotomous items it is called Item Response Theory]).

Discussion
The impact of robust statistical methods to provide construct validity is still limited, being majority of psychological tests still were based on classical test theory. This is clearly an important limitation. Normally the proposed scales, tests, and batteries uses Cronbach's alpha as a reliability index; however, such index is only a measurement of inter-item covariance per unit composite variance and a reasonably good index of reliability under certain restrictive condition which, in a brief review of the above cited tested, have been never assessed to ensure Cronbach's Alpha's use. Robust statistical models must be applied to music perception research, providing better understanding how to generate reliable, viable, and precise phenotypical measures.

Information dynamics of boundary perception: Entropy in self-paced music listening

Haley E. Kragness[1], Niels Chr. Hansen[2], Peter Vuust[2], Laurel J. Trainor[1,3,4], Marcus T. Pearce[5]
[1]Psychology, Neuroscience and Behaviour, McMaster University, Hamilton, ON, Canada
[2]Center for Music in the Brain, Department of Clinical Medicine, Aarhus University & The Royal Academy of Music Aarhus/Aalborg, Denmark
[3]McMaster Institute for Music and the Mind, McMaster University, Hamilton, ON, Canada
[4]Rotman Research Institute, Baycrest Hospital, Toronto, Canada
[5]Cognitive Science Research Group and Centre for Digital Music, Queen Mary, University of London, UK

It has long been noted that expert musicians lengthen notes at phrase boundaries in expressive performance. Recently, we have extended research on this phenomenon by showing that undergraduates with no formal musical training and children as young as 3 years lengthen phrase boundaries during self-paced listening to chord sequences in a lab setting (the *musical dwell-time effect*). However, the origin of the musical dwell-time effect is still unknown. Recent work has demonstrated that musicians and non-musicians are sensitive to entropy in musical sequences, experiencing high-entropy contexts as more uncertain than low-entropy contexts. Because phrase boundaries tend to afford high-entropy continuations, thus generating uncertain expectations in the listener, one possibility is that boundary perception is directly related to entropy. In other words, it may be hypothesized that entropy underlies the musical dwell-time effect rather than boundary status per se. The current experiment thus investigates the contributions of boundary status and predictive uncertainty to the musical dwell time effect by controlling these usually highly-correlated factors independently. In this procedure, participants self-pace through short melodies (derived from a corpus of Bach chorales) using a computer key to control the onset of each successive note with the explicit goal of optimizing their memorization of the music. Each melody contains a target note that (1) is phrase ending or beginning and (2) has high or low entropy (as estimated by the Information Dynamics of Music Model, IDyOM, trained on a large corpus of hymns and folksongs). Data collection is ongoing. The main analysis will examine whether longer dwelling is associated with boundary status or entropy. Results from this study will extend recent work on predictive uncertainty to the timing domain, as well as potentially answer key questions relating to boundary perception in musical listening.

We are fast learners: the implicit learning of a novel music system from a short exposure

Yvonne Leung[1], Roger T. Dean[1]
[1]MARCS Institute, Western Sydney University, Sydney, NSW, Australia

We examined if one can incidentally learn a microtonal scale through listening to microtonal melodies during a timbre discrimination task, where participants are asked to detect a timbre change in one of the notes (the target) in a melody. Utilizing a novel approach, we manipulated the note before the target (the prime) such that it was either from the musical scale of the presented melody, or from another. Unlike past paradigms where timbre change was manipulated on the target, this paradigm avoided having both the pitch manipulation and timbre alteration on the same note. When the prime is from a different scale, it will create two incongruent pitch intervals with the notes before and after it. We hypothesized that if participants perceive those pitch intervals as incongruent, they would respond faster to the target, as a timbre change after the incongruent intervals becomes highly expected due to us providing such an increased probabilistic association in the design.

Melodies in the task were blocked by scale: a diatonic scale (Control) or a microtonal scale. To provide exposure, participants first performed a 10-minute discrimination task where no incongruent intervals were embedded in the melodies. They were then tested using the new paradigm immediately and again a week later. Results from 21 non-musicians showed that immediately after the exposure, response time was significantly shorter when targets were preceded by incongruent intervals than when they were not, for both the Control and Microtonal melodies. However, this difference was only significant for the Control melodies a week later, suggesting that one short exposure session is not sufficient for long-term memory encoding. Nevertheless, this consistent pattern found in the Control stimuli suggests that this new paradigm is effective in measuring implicit knowledge of music, in this case, listeners' consistent ability to differentiate incongruent pitch intervals from the intervals in a diatonic scale.

Tonal cognition: a dynamic approach

Duilio D'Alfonso

***Abstract*—**Tonal cognition can be studied either in a probabilistic perspective or in a deterministic view. I will argue that in the former case what is characterized is a "weak tonal sense", while the "strong tonal sense" must be tackled by means of deterministic, symbolic, rule-based approaches. In this framework, I propose to transfer in music cognition the "dynamic" perspective developed, in the last decades, in logic and linguistics. With this purpose, I present a dynamic model of tonal cognition, essentially based on the generative theory of tonality of Lerdahl and Jackendoff, modified to incorporate dynamic principles in the real-time process of events hierarchy inference.

***Keywords*—** Tonality, Prolongational Theory, Language/Music Parsing, Dynamic Syntax.

I. Introduction

TONAL cognition, the listener's ability to develop the sense of the key of a tonal piece in the listening process, is connected to at least two structural aspects of tonal Western music: a frequency distribution on the occurrences of pitch–class sets with a prevalence of tones belonging to a major/minor scale (according to Krumhansl's "key–profiles" [1]); a hierarchical organization of (ordered) pitch–events defining tensing/relaxing patterns. Truly, a role is played also by the recognition of linear patterns in voice leading, at least those involving the main voice, the *cantus*, and the bass (as pointed out by Schenkerian back/middle–ground graphs), but in this paper I intend to focus on the "harmonic" structural patterns determining tonal cognition. The unfolding of chords progressions (triads and seventh chords as the harmonic alphabets of tonal Western harmony), also when chords are linearized, is the infrastructural basis of the tonal sense, while it is not clear if linear patterns can be considered, in some sense, infrastructural as well (see Temperley [2] for what is intended as infrastructural aspects of music).

Now, we can distinguish a "weak" tonal sense, due to distribution and density of pcs events, widely studied by probabilistic models of key–finding, and a "strong" tonal apprehension, represented by event hierarchies recovered by experienced listeners, that has received less attention, especially regarding the cognitive and computational task to recover it in the real-time listening process.

Thus, in this paper I advance a model of the "event hierarchy inference", a "left-to-right" incremental parsing strategy to derive the event hierarchy in the format of prolongational trees. A locality principle to constrain the accessibility of the tree under construction is introduce to account for listeners limitations in memory. Preference rules for prolongational reduction, in the Lerdahl and Jackendoff grammar of tonal music, the well-known Generative Theory of Tonal Music (GTTM), are hardly conceivable as a real-time parsing system [3]. Their system licenses the best prolongational descriptions derivable by an idealized listener (as pointed out by the same Jackendoff [4]): such descriptions could not coincide with those inferred by finite users. Especially for large and complex pieces, it is evident that the structural descriptions realistically inferred by "normal" listeners could not be coincident with those theoretically provided by GTTM. Therefore, in my proposal, "affordable" descriptions, derivable by resource–bounded listeners, are allowed, even if divergent from the most sensible ones.

Moreover, as in Dynamic Syntax (Kempson *et al.* [5]), where natural language grammar is conceived as a goal–oriented real–time parsing system, the musical parsing process is guided by the initial purpose to recover a basic tree structure (similar to the normative tree structure of Lerdahl and Jackendoff), and proceeds alternating bottom-up input scanning and top-down prediction steps. The idea is that listener's mind is neither a *tabula rasa* before listening nor passive during listening; instead, it is in the active state to predict the tonal sense and, triggered by the ongoing musical stimuli, it generates goals and sub-goals orienting the comprehension process.

II. The weak tonal sense

Actually, what I call the weak tonal sense is (relatively) independent from the melodic or harmonic order of pitch events. Just to give a simple idea of this point, consider a music sample such as the first measure of the Bach's Prelude in C minor (from the 2nd volume of the Well-Tempered Clavier):

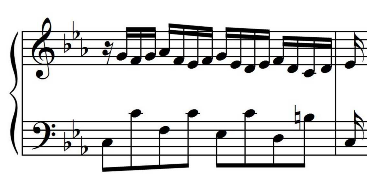


Fig. 1 The first measure of the Bach's Prelude in C minor, WTC, 2nd volume.

Here, the strong sense of C minor is mainly given by the chords order (i–iv–i^6–vii^{o6}–i), corroborated by the contrapuntal pattern of the two external voices, as emphasized in fig. 2.

Duilio D'Alfonso (PhD in Philosophy of Language) is professor of Music Theory and Analysis at the Music Conservatory "Ottorino Respighi", Via Ezio 32, 04100 Latina, Italy, www.conslatina.it (corresponding author phone: 06-44244320; mobile phone 0039-3287691872; e-mail: duilio.dalfonso@tin.it).

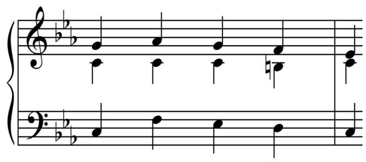


Fig. 2 The "reduction" of the previous example.

The closing chord is strongly affirmed as the tonal center, representing the more stable harmonic event. But, imagine a random permutation of the notes, preserving pitches and durations, so that the frequency distribution among pitches is maintained, as in the following example:

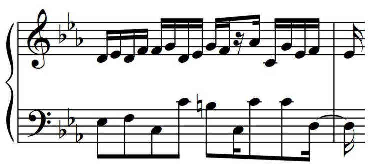



Fig. 3 A permutation of the notes (and pauses) of the example in fig. 1, preserving the occurrences of tones and the temporal density of events.

The intuition, here, is that the tonal sense of C minor is not completely vanished; rather, it has been deteriorated by the loss of the cadential meaning of the chords progression in the original passage, as well as, maybe, as a consequence of the disappearance of the upper voice linear (descending) pattern, 5–6–5–4–3, forming with the bass line 1–4–3–2–1 a contrapuntal pattern of descending parallel tenths. Straightforwardly, we can say that these two musical passages receive the same score by the Krumhansl key–finding algorithm, and thus the same key; notwithstanding, they are very different in the sense of tonality they induce in listeners.

Hence, we can hypothesize a weak tonal sense, essentially due to distribution and density of pitches, the order of which is irrelevant, as what is theoretically investigated by probabilistic models of key–finding, in the sub–symbolic perspective on music cognition, also inspired by the paradigm of neural computing (see Tillmann *et al.* [6]). Empirically, the weak tonal cognition has been studied by setting up experiments based on the probe-tone methodology (a number of such experiments are reported in literature; for the dynamics of tonality perception see Toiviainen *et al.* [7]).

III. THE STRONG TONAL SENSE

The strong tonal sense is related, in the essence, to two aspects of the musical experience: tensing/relaxing patterns and expectation/satisfaction patterns (following Narmour [8], we could speak of implication/realization patterns, since the expectation is determined by an implication). Both aspects, strictly connected, are maximally exploited in, but not exclusive of, tonal Western music. These two experiential features of the listening to music are intensely related to syntactic features of music, and particularly to the order of events. As such, they are hard to be investigated within the probabilistic methodologies, and probabilistic models dealing with orders of pitch–events are necessarily based on wide simplifications. Typically, such models are inspired to "Markov chains", so that each event is conditioned only by the previous one. But, obviously, the process of music understanding is not (only) a Markovian process. Limited to melody perception, a Markovian approach has been showed to be very useful, especially to account for the familiarity of melodic patterns (see Temperley [9]). Despite the important results of the statistical approaches to tonal cognition (for the relevance of probability in music see, also, Meyer[1] [10]), my point here is that there is a strong tonal sense inferred by experienced listeners that cannot be dealt with by means of probabilistic methods; we need rule-based symbolic approaches in order to address tonal cognition.

Now, in the last decades a new paradigm, in the general framework of formal, deterministic, and rule–based perspective on logic, language and information, has emerged: the paradigm of "grammar as action". This paradigm was inspired by the rise of dynamic logic and semantics (van Benthem [11]), and successively extended to other areas of cognition and applied to linguistic theory, as "dynamic grammar", by Kempson *et al.* [5], as well as in discourse analysis by Chafe [12], even if from a different viewpoint[2]. According to the dynamic grammar view, language grammar is a parsing strategy, finalized to recover meaning in a real–time, sequential, incremental processing of the linguistic input, i.e. strings of words. The parsing process proceeds step–by–step, each step taking the current word and the structural context determined by previous ones as input, and releasing an update of that context. The process in goal–oriented: the goal is to establish some propositional content, and the initial goal is divided into sub–goals, according to the information progressively available.

The parallel with music is natural: music understanding seems to involve time–linear sequential processing, controlled by real–time scanning of the input and top–down expectations generated by hierarchically organized structures. The main difference between music and language is the absence of propositional content in music, that is the principal goal governing the language parsing strategy. But, as I will argue, since tonal hearing is a goal–oriented process, for tonal music, the main goal guiding the comprehension process is the recovery of the strong tonal sense, represented by the so-called "basic normative structure" in the GTTM.

With this purpose, I will sketch a model of musical parsing strategy, relatively to the "tonal content" of music (with meter and grouping structure in background), that is an attempt to convert the GTTM into a dynamic grammar of tonal music: in a motto, from generative theory of tonal music to dynamic theory of tonal music. In such a perspective, since action–based architecture is constitutive of the musical grammar, we can account for listeners limitations (in attention, working–memory, etc.) in a more natural fashion. Synthetically, what I suggest is that the shift from classical, static and "infinitary" conceptions towards dynamic and "finitary" perspectives in

[1] I like to report here the following consideration on the role of statistics in the analysis of musical style: "Since all classifications and all generalization about styles are based on some estimate of relative frequency, statistics are inescapable. This being so, it seems prudent to gather, analyze, and interpret statistical data according to some coherent, even systematic plan." Meyer [10].

[2] For dynamic principle in discourse theory see also Kamp *et al.* [13] and Peregrin [14].

logic and information theory could be fruitful also in music cognition, and especially in tonal cognition.

IV. The Stability Theory

It could be useful to spend some words on the problem of the so-called stability theory. According to Lerdahl and Jackendoff, a stability theory provides a (formal) criterion to assign stability value to pitch–events, distinguishing the more stable events from the less ones. In Lerdahl [15], a tonal space is introduced, with the relative metric to determine the distance between chords, that is mainly thought with the purpose to account for the tonal context dependence of the perceived tension, or distance, between chords. Stability values of chords are determined on the basis of an economy principle: given a string of chords, the “tonal reading” of the entire string that minimizes distances (or, better, sets to the smallest values the tonal tension) is preferred.

Now, I adopt a similar view, but under the assumption that listeners attempt to recover the tonal sense, rather than the minimal path in the tonal space, in the format of what Lerdahl and Jackendoff call the “basic normative structure”, showed in fig. 4.

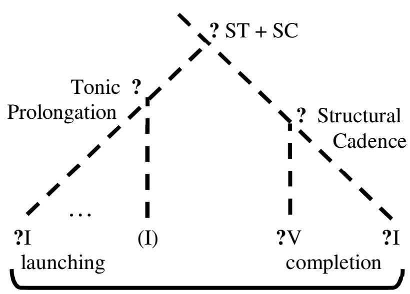


Fig. 4 The basic normative structure, in the form of prolongational tree. The main right branch covers the initial structural tonic prolongation. It could imply multiple reaffirmation of the tonic chord. The main left branch dominates the closing structural cadence. The dashed tree is intended to express that the tree is a prediction of the listener, a kind of dispositional state to be accomplished. The horizontal square bracket is intended to suggest that the basic tree is generally confined within a phrase of certain length, a phrase that it contributes to determine. For the interaction between dashed branch and question marks, see section 5.

Thus, the stability theory, in a sense, can be incorporated in the parsing system rather than formulated as a separated module of the grammar, according to the postulate that listeners’ task is to map the structure in fig. 4 into the string of events they perceive.

For the sake of exemplification, let take the chords chain G–C (both major). We know that it can be interpreted in both G major (I–IV) or C major (V–I) key. But, as shown in fig. 5, supposing that no event precedes the chain, the listener’s attempting to apply the basic structure successes only by attaining the chain to the right branch of the expected tree. This prediction will then be reinforced if the chain is followed by a cadence in G major. Obviously, it could be also refuted: it should be considered a kind of “default reasoning”, always revisable if new information contradicts it. Anyway, if the string G–C is followed by a progression confirming the C major key, it can persist, at least for few time, the memory of the right branch initially unsatisfied. This is a way to account for the sense of “unusualness” of a piece in C major starting with the dominant chord.

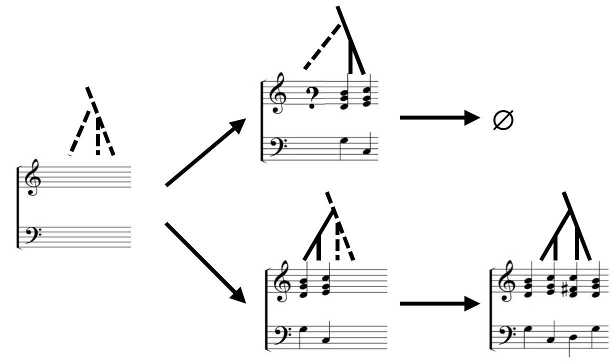


Fig. 5 Assuming the basic normative structure as the initial predictive state of the parsing process, once perceived the progression G-C, the listener hypothesizes two possible interpretation, represented by the two staves in column. But, only the lower one, implying the prediction of a structural cadence, can be satisfied, while the upper one is predestined to failure.

Similar arguments can be used to justify the examples in fig. 6. Prolongational descriptions of the chords progressions are the result of the application of the prolongational structure in fig. 4 to the available input. The trees above the staves should be intended as representation of the listeners’ mental states. In the upper staff the listener, once received a chord progression fulfilling the expectation of tonic prolongation, predicts, and waits for, a structural cadence, possibly introduced by subdominant chords. In the lower staff, the listener’s expectation, once processed a perfect authentic cadence (PAC), is (tentatively) fulfilled.

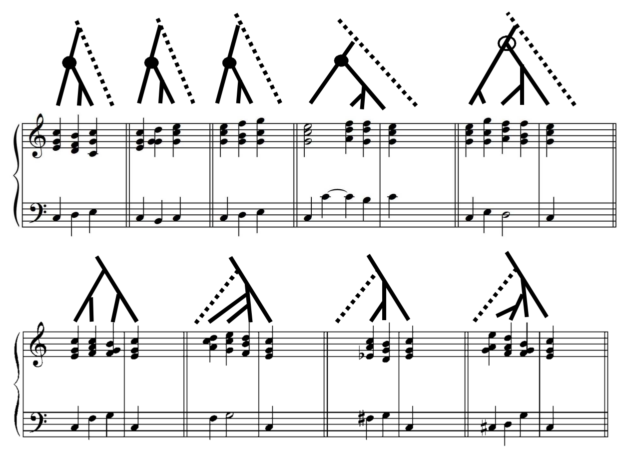


Fig. 6 In the upper staff are listed some basic chords progressions prolonging the structural tonic; in the lower staff are shown some cadential progressions completing the basic normative structure. In the last three case the rightmost branch dominates the structural beginning of the progressions, evidently a tonic prolongation.

The examples in fig. 6 show a kind of “locality principle” at work: given a string of two chords, the less stable is attached to the more stable; for three chords, prefer recursive structures; do not license recursive trees with a single dominant branch covering string longer than four chords.

Approximately, we can consider local context not longer than involving four chords. Generally, dominant sevenths and (half)–diminished chords (both triads and sevenths) are attached to their following chord, as their "resolution"[3].

V. THE REAL–TIME PARSING SYSTEM

Probably, the only attempt to characterize the real–time process of the comprehension of a tonal piece has been carried on by Jackendoff [4], where he describes the incremental growing of mental representations in a listener who listens to a Bach's choral first measures. He focused especially on the representation of the meter.

In my proposal, that is a work in progress, the meter is not directly addressed, and somewhat presupposed. I assume also the phrase-structure description, as interacting with the growth of the prolongational tree. Note also that the distinction between time–span reduction and prolongational reduction, as formulated in the GTTM, is not necessary, since the construction of the prolongational tree is not a top–down procedure receiving in input a whole time–span reduction. Such process is performed as a bottom–up parsing, constrained by top–down predictions (similar to what is known as "chart parsing" in the context–free languages parsing literature, see [16]). It proceeds alternating bottom–up input scanning and top–down prediction steps, and it is limited to phrases of what we can call the "sentential level". For example, assuming Caplin terminology [17] for the music of the classical style, we can suppose that a normative structure is unfolded by the 8–measure typical theme, in the form of *sentence* or *period*, as well as by a section of the *small ternary* form (exposition, recapitulation). More problematic is to deal with the larger forms, such as the Sonata–form. Consider only how difficult is to establish wheatear or not listeners are able to grasp the tonal closure of long pieces [18]. So, at the moment, I consider a unique prolongational tree for large–scale forms more as a theoretical construct than a representation with cognitive plausibility. Its cognitive relevance should be further investigated.

Let me now briefly address the real–time process of the growth of the prolongational tree. The parsing procedure begins with a prediction step, by predicting the basic normative structure in fig. 4. This corresponds to the initial listener's expectation, or disposition, to derive a unique, hierarchical, and coherent tonal representation from the ongoing auditory stimulus. The listener tries to fulfill the requirements determined by this expectation, represented by the question marks labeling the nodes of the tree. It may be worth to note that the tree in fig. 4 is similar to that showed by Lerdahl in [19] (displayed in fig. 7); the Lerdahl tree, representing the basic structural organization of the tension/relaxation pattern, can be considered a more general characterization of the tree in fig. 4, the latter being an adaptation of the former to the tonal, classical, idiom.

[3] Really, seventh chords have always been quite problematic. They have never been inserted in a tonal space. Just to say, neither the Lerdahl tonal space nor the Riemannian *Tonnetz* contemplate the seventh chords. Adding a seventh to a triad means to decrease its stability, both because sevenths are dissonance (at least in the classical tradition) and because they originate as passing tones. This implies that the stability of the subsequent chord is somewhat emphasized.

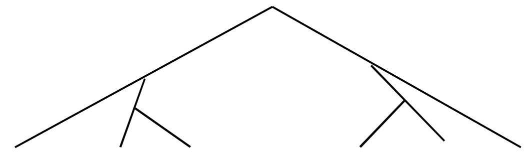


Fig. 7 The typical ordering of tonal tensions, according to Lerdahl.

Now, we must distinguish between (un)satisfied and (un)saturated branches and trees (here and henceforth, with "tree" I mean every sub-tree of the main tree – rooted in a node dominated by the root of the main tree, as well as the main tree itself). As said, question marks are the representation of requirements. A requirement decorating a node t, the root of a tree T, is satisfied if all the requirements in every proper sub-tree of T are satisfied; otherwise, the requirement is unsatisfied. A tree is satisfied if the requirement in its root is satisfied. A question mark decorating a node t can be deleted if the tree rooted in t is satisfied.

A branch of the starting dashed tree becomes solid if it points to a pitch-event. In such a case we say that it is saturated, unsaturated otherwise. We say that the starting dashed tree is unsaturated, trees with some, but not all, solid branches, produced in the course of the parsing process, are partially saturated. A tree T is called saturated if all branches of T are solid; this definition implies that a tree T is satisfied if T is a saturated tree with all the question marks deleted.

We must allow for solid trees to be unsatisfied and partially saturated trees to be satisfied, as we shall see in what follows; the above definitions assure this feature, since satisfaction and saturation are independent and independently defined. Naturally, although formal definitions of requirement satisfaction would be desirable, here we informally say that the question mark decorating the higher node on the left branch of the starting tree is satisfied if the tree rooted in this node covers the initial tonic prolongation; the question mark decorating the higher node on the right branch is satisfied if it covers the closing (perfect authentic) cadence. The left branch departing from the node covering the cadence is satisfied if saturated by the dominant chord (possibly, in root position); the corresponding right branch is satisfied if saturated by the tonic chord, in root position. Note that, in the starting tree (fig. 4), only the leftmost leaf dominated by the node covering the tonic prolongation is decorated with a question mark. This is because the tonic prolongation could consist of one or more unfolded tonic chords – even just the first tonic can satisfy the tonic prolongation requirement. The final successful state of the parsing is given by the satisfied tree covering the whole input string.

An oversimplified example of the process is shown in fig. 8. In 1) is represented the starting configuration of the parsing process: the input scanning is waiting to start and the tree is unsaturated. In 2), received the first two chords, the system projects a partially unsaturated tree: in fact, the progression I–I^6 saturates the right branch and another one attached to it. The topmost node question mark cannot be deleted, as the left branch of this node isn't still become solid, what happens when the last two chords are processed.

Note that the tree in 2) is to be intended as a partial representation. As for dynamic grammars for natural language, musical parsing should be able to deal with partial information, if we want it to manage the dynamical, and contextual, aspect of mind (musical) information processing.

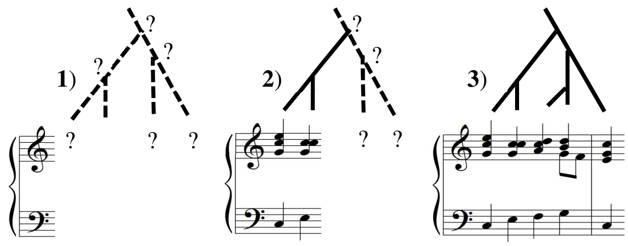


Fig. 8 From the left to the right: three state of the chords progression parsing, represented by the trees above the staves.

In this simple case, the system performs a top-down prediction step as the initialization of the parsing process; in the course of the process only bottom-up steps were necessary, input scanning followed by bottom-up (depth-first) parsing steps, to integrate the input in the expected structure. But, of course, the system should be endowed with the capacity of manipulating the tree under construction, triggered by the acquired information, but without performing a scanning step, i.e. the capacity of performing top-down prediction steps during the parsing process, eventually projecting sub-goals of the main goal to fulfill the question mark in the basic tree root. A number of such cases should be provided, with the additional difficulty to obtain a computable system aiming at the completeness. Here, I'll illustrate three case, just in order the give an idea of this device.

First case. Consider the first measure of the Bach's choral, in A minor shown in fig. 9. It begins with the dominant chord resolving to the tonic followed by another dominant-tonic pattern ($vii^{\circ 6}$–i^{6}). What is here the parsing strategy of the hearer? Probably, she postulates a hidden initial tonic when she grasps the sense of A minor; but this postulate cannot be stored for long time. To account for this fact, we can provide the system with a rule transforming a tree with a rightmost unsaturated branch into a tree with a backward prolongation of the main right branch. This rule involves, evidently, a pure prediction step, forced by a situation where a highly unstable representation relaxes into a more stable one.

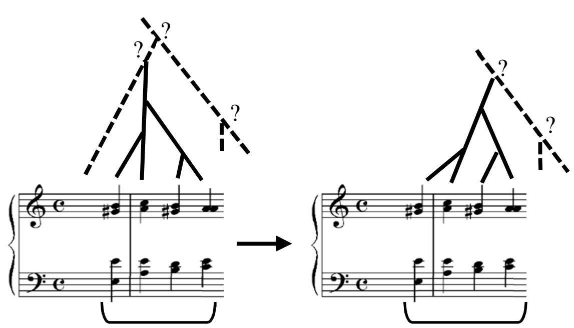


Fig. 9 The first four chords of the Bach's choral *Ach Gott, vom Himmel sieh' darein*, Cantata 153.

Second case. A listener, after having saturated the basic tree with the final tonic chord replaced by, say, the vi chord, as in the deceptive cadence, feels that the requirement of the closing cadence (i.e., the question mark decorating the higher node in the right branch) is unsatisfied. Thus, also the question mark in the root of the basic tree remains undeleted, although the branches of the tree are solid (the tree is saturated). This situation can be expressed by the tree 1) in fig. 10. As you can see, the tree is saturated but unsatisfied, as the question marks along the branch pointing to the closing tonic are still in the waiting for satisfaction. This unstable representation induces the projection of a new, unsaturated right branch, and this is, similarly to the previous case, a pure prediction step. The listener is now expecting a new cadence, as represented by the tree 2), since the cadence just processed is not able to satisfy the expectation of closure generated by the tonic prolongation followed by subdominant chords previously perceived.

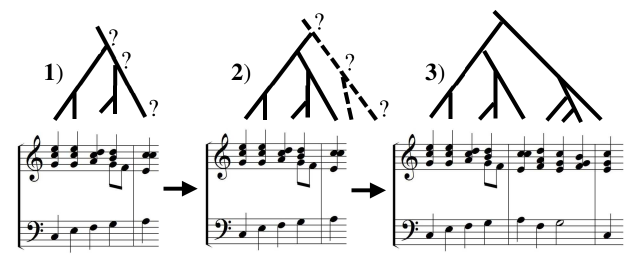


Fig. 10 The processing of the deceptive cadence. A pure prediction move induces the projection of a novel right, cadential, branch, as the deceptive cadence produces a sense of failure of the basic structure accomplishment.

Third case. The well-known structure of the 8-measure theme in the classical style, in the form of period, is characterized by the so-called *interruption*[4], at the end of the first 4 bars (the *antecedent*). The interruption implies a half–cadence, with the V positioned immediately before the border of the antecedent phrase. Again, how does the listener build up a coherent representation of the interruption, and how does she merge this representation with that of the *consequent* (the second 4–measure of the period, usually concluded by a PAC)? My hypothesis is that the interruption implies a sense of incompleteness, but it is such that, in order to receive accomplishment, it generates the expectation of a re-launch of the entire normative structure. The sense of interruption is emphasized also by means of other parameters, marking the phrase border, independently from the harmony (usually there is a kind of textural caesura after the dominant chord, at the end of the antecedent). This situation is represented in fig. 11. In 1), the listener temporary remains on hold for completing the cadence. Then, triggered by the perception of the phrase border, she projects a new entire unsaturated tree as the expectation of the accomplishment of the interruption, as shown in 2).

[4] The term is imported from the Schenkerian tradition in music analysis. It refers to the stop at the supertonic (the second scale-degree) in the descent of the fundamental linear progression of the fundamental structure (the *Ursatz*), with the bass rising, and stopping, at the V. This produces a half-cadence, that is considered an interruption in the harmonic flow. After the interruption, the *Ursatz* is entirely replicated (see [20] for more details).

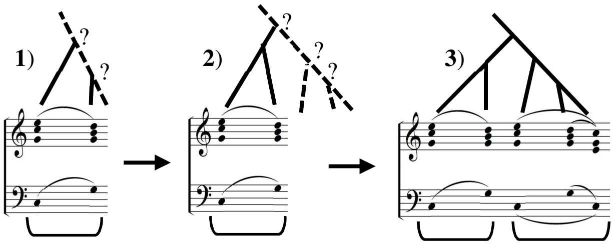


Fig. 11 An hypothesis of the real-time processing of the interruption, after the first 4 bars, in the eight-measure theme known as period.

VI. Conclusion

The model sketched is informal and requires to be formally outlined. It is a working hypothesis to be developed and tested. It has to be formalized in order to become a model of the listener tonal competence experimentally testable and it is surely desirable it achieves some kind of completeness. But completeness in music theory is not the same as in logic. The dynamic grammar of tonal music should be a flexible system, incorporating the principles of dynamic processing of information, already adopted in many non-classical logics, as well as in linguistics. More than theorems of completeness, in theoretical psychology of music we need of formal hypotheses that can be empirically testable and that, at the same time, can account for the cognitive complexity of the musical experience.

I argued for the necessity to introduce a way of representing "partiality" in music processing: the complex interaction between perception and prediction, in a non-monotonic growth of representations, is a crucial aspect of music understanding, and in order to deal with this aspect we need of models managing incomplete structures. Anyway, it is already a long time that formal systems and logics dealing with partiality have been introduced in theoretical linguistics.

To sum up, my principal purpose here has been to argue for the fruitful adoption in music cognitive theory of the dynamic, symbolic perspective, as a paradigm developed in logic and semantics but with relevant involvement in mental information processing. Tonal cognition, as a special case of music cognition, shows some features that seem particularly suitable to be treated in this perspective, such as a strong context dependence in the information processing and a strong tendency of the mind to actively interact with perception. The current information in perception is interpreted against a structural context produced by the data already processed. Moreover, tonal hearing amplifies a pervasive character of the listening to music: that to project the past in the future. What is in memory and what is currently perceived prompt the hearer to predict what is going to be. Now, exactly these aspects of the musical experience are efficiently captured by the dynamic perspective, to the extent that dynamic principles have been introduced just to address two main features characterizing mental understanding processes in human cognition: the general context dependence (among others, involving temporal constraints) and the tendency to anticipate the future.

References

[1] C. L. Krumhansl, *Cognitive Foundations of Musical Pitch*, New York: Oxford University Press, 1990.

[2] D. Temperley, *The Cognition of Basic Musical Structures*, MA: MIT Press, 2001.

[3] F. Lerdahl and R. Jackendoff, *A Generative Theory of Tonal Music*, MA: The MIT Press, 1983.

[4] R. Jackendoff, "Musical Parsing and Musical Affects," *Music Perception*, vol. 9, No. 2, pp. 199-230, 1991.

[5] R. Kempson, W. Meyer-Viol, and D. Gabbay, *Dynamic Syntax: the flow of Language Understanding*, UK: Blackwell, 2001.

[6] B. Tillmann, J.J. Bharucha, E. Bigand, "Implicit learning of tonality: A self-organizing approach." *Psychological Review*, vol. 107(4), pp. 885-913, 2000.

[7] P. Toiviainen and C. L. Krumhansl, "Measuring and modeling real-time responses to music: The dynamics of tonality induction", *Perception*, vol. 32, pp. 741-766, 2003.

[8] E. Narmour, *The Analysis and Cognition of Basic Melodic Structures: the Implication-Realization Model*, Chicago: The University of Chicago Press, 1990.

[9] D. Temperley, *Music and Probability*, MA: The MIT Press, 2007.

[10] L. B. Meyer, *Style and Music: Theory, History and Ideology*, Chicago: The University of Chicago Press, 1989.

[11] J. van Benthem, *Language in Action*, The MIT Press, 1995.

[12] W. Chafe, *Discourse, Consciousness, and time*, Chicago: The University of Chicago Press, 1994.

[13] H. Kamp and U. Reyle, *From discourse to Logic*. Dordrecht: Kluwer, 1993.

[14] J. Peregrin (ed.), *Meaning: the Dynamic Turn*, Oxford: Elsevier Science, 2003.

[15] F. Lerdahl, *Tonal Pitch, Space*, Oxford University Press, 2004.

[16] D. Jurafsky and J.H. Martin, *Speech and language processing*, New Jersey: Prentice-Hall, 2000.

[17] W.E. Caplin, *Classical Form*, Oxford: Oxford University Press, 1998.

[18] N. Cook, "The Perception of Large-Scale Tonal Closure," *Music Perception*, vol. 5 pp. 197-205, 1987.

[19] F. Lerdahl, "Musical Syntax and its relation with Linguistic Syntax", in M.A. Arbib (ed.), *Language, Music, and the Brain*, MA: The MIT Press, 2013.

[20] D. Gagné and A. Cadwallader, *Analysis of Tonal Music*, Oxford: Oxford University Press, 2007.

On the Role of Semitone Intervals in Melodic Organization: Yearning vs. Baby Steps

Hubert Léveillé Gauvin,[*1] David Huron,[*2] Daniel Shanahan[#3]

[*] *School of Music, Ohio State University, USA*
[#] *School of Music, Louisiana State University, USA*
[1]leveillegauvin.1@osu.edu, [2]huron.1@osu.edu, [3]dshanahan@lsu.edu

ABSTRACT

A corpus study of 15,000 melodies was carried out to test various conjectures arising from the purported tending or yearning quality of the semitone interval. Several hypotheses were tested, including comparing semitone and whole-tone intervals in major-mode contexts with comparable movements in minor-mode contexts. Results failed to show the predicted relationship favoring semitone intervals for "yearning."

In the course of this study, our analyses produced an unexpected result regarding the use of semitone intervals in melodies generally. Although whole-tone intervals are more common than semitone intervals, compared with melodies whose pitches have been randomly reordered, there is a tendency for composers to favor semitone intervals compared with whole-tone intervals. These results reinforce longstanding observations regarding pitch-proximity, but add to these observations by identifying a disposition toward using the smallest available intervals in the construction of melodies. Overall, the results call into question the conjecture that the semitone interval itself is responsible for the yearning qualia.

I. INTRODUCTION

The strong affinity between certain pitch relationships has been the topic of research by both music theorists and music psychologists. Among music theorists, pitches related by a semitone interval have attracted considerable speculation and discussion. Specifically, the semitone is commonly regarded as exhibiting inherently unstable properties. In *Ars Musica* (1296-1304), Aegidius of Zamora referred to the semitone as an "imperfect tone" that must resolve in order to reach perfection (cited in Leach, 2006, and Fuller, 2011). Boethius similarly wrote that the very term "semitone" was derived from *semus, sema, semum,* meaning imperfect (cited in Fuller, 2011). This notion of an "imperfect," unstable semitone has been consistent throughout the history of music theory. Fétis (1844) believed that tonality was itself a consequence of the attractions from scale degree $\hat{7}$ to $\hat{1}$, and from scale degree $\hat{4}$ to $\hat{3}$. Similarly, Lussy (1874) wrote that tones exhibit higher levels of attraction depending on their relative proximity. More recently, Lerdahl (2001, 167) discussed at length "the psychological need for an unstable pitch to be assimilated to an immediately subsequent proximate and stable pitch."

Although the purportedly unstable semitone has been a central topic in music theory for centuries, it has received empirical attention only in recent decades. Krumhansl and Shepard (1979) empirically demonstrated a hierarchy of stability for different scale tones, and showed that there is a psychological tendency for less stable tones to proceed to more stable tones. Similarly, Deutsch (1978) showed the importance of pitch proximity in processing tones. Bharucha (1984, 1996) combined the properties of both proximity and stability in his model of melodic anchoring, stating that "two constraints characterize the process of melodic anchoring: the anchor and the anchored tones are close in pitch (proximity), and the anchor always follows the anchored tone (asymmetry)" (1996, 383). Bharucha refers to the directional expectation generated by both of these factors as the "yearning vector," which he defined as "the psychological force pulling the [musical] event up or down" (1996, 393). He carried out a behavioral study that demonstrated that listeners perceived the distance from scale degree $\hat{7}$ to scale degree $\hat{1}$ as closer than the reverse distance of $\hat{1}$ to $\hat{7}$ (1984). More generally, listeners perceive an inherent proximity when an unstable pitch resolves to a stable one, but not the other way around. Similarly, Francès demonstrated that listeners are more likely to detect when a leading tone is too low rather than too high (1958/1988, discussed in Yeary, 2011).

At least in the case of Western music, musicians commonly refer to the "yearning" qualia of a semitone relationship. Consider the semitone relationship between $\hat{7}$ and $\hat{1}$. For Western-enculturated listeners, the "leading tone" is aptly named as it evokes a strong urging or longing feeling. But this phenomenon is not limited to the upward semitone pitch movement from $\hat{7}$-$\hat{1}$. In certain musical circumstances the same feeling of yearning attends the movement between the fourth and third scale degrees. A related example can be found in the so-called $\hat{4}$-$\hat{3}$ suspension, in which a downward semitone tendency is delayed—heightening the phenomenological experience of yearning. The qualia associated with different scale tones have been chronicled in studies by Huron (2006) and Arthur (2015) in which musician listeners were asked to describe the feelings evoked by various scale degrees. The leading tone was described using terms such as *inevitability, unstable, pointing, restless, uncomfortable, squirmy,* and *itching.* In analyzing the content of such opened-end descriptions, Huron distinguished a semantic category dubbed "stability," and linked it to first-order probabilities in exposure. That is, when stimulus X is commonly followed by stimulus Y in the environment of exposure, then when listeners encounter stimulus X, they commonly experience a strong feeling of instability and anticipation of the ensuing occurrence of Y. However, Huron's study focused on the qualia of scale tones and did not examine qualia related specifically to intervals. In other words, Huron's 2006 discussion did not link yearning qualia in any way to the interval of the semitone.

A number of theoretical ideas have been proposed to explain the possible origins for what might be dubbed "the yearning semitone." For example, it is possible that the operative principle is that when you are close to something important, central or expected, there is a strong sense that one should move to that something. An alternative view might be that there is nothing inherent to small intervals that would lead to this yearning quality. It may simply be that scale degrees $\hat{7}$ and $\hat{1}$ have a strong attachment, and that scale degrees $\hat{3}$ and $\hat{4}$ have similarly strong attachments independent of the semitone relationship. The implicit learning view posits no special function for the semitone: any pitch might accrue yearning qualia simply by its statistical tendency to be followed by some other pitch. For example, the dominant might evoke feelings of anticipation for the tonic, simply because in Western music, many dominants are followed by the tonic.

These speculations aside, the purpose of the present study is not to resolve the issue of origins, but to empirically test the existence of the phenomenon. That is, our aim is to determine whether musical organization is indeed consistent with the special role of the semitone in such a yearning or tending relationship. Accordingly, we might propose the following conjecture:

Scale degrees separated by a semitone are more likely to cleave together than scale degrees separated by a larger distance (e.g., whole tone).

In testing this conjecture we face at least two potential confounds. The first difficulty relates to compositional intent. Without resorting to a perceptual experiment, how might we operationalize the notion of "cleaving" or "yearning"? If one tone tends to cleave to another tone, evidence consistent with this relationship might minimally involve a statistical tendency for the one tone to be followed by the other tone. Of course, in real music, composers might aim to increase tension or engage in deception by interposing a third tone between a purported "yearning" pitch and a purported "yearned-for" pitch. So a simple tally of the number of X followed by Y may be necessary but not sufficient evidence consistent with a purported "cleaving" or "yearning" qualia. Nevertheless, we may reasonably suppose that evidence of a purported "cleaving" quale would minimally involve an elevated likelihood that one of the tones will have a high probability of being followed by the other tone.

A second potential confound relates to the relative stability of different scale tones. Musical melodies are not simply successions of intervals. Melodies are also salient successions of scale degrees, and some scale tones are more important than others. For example, scale degrees $\hat{1}$, $\hat{3}$ and $\hat{5}$ are known to be more stable than other scale tones. These differences are empirically evident, for example, in the key profiles assembled by Krumhansl and Kessler (1982). Unstable tones are simply attracted to more stable tones. This suggests that the tendency for $\hat{7}$ to move to $\hat{1}$, and for $\hat{4}$ to move to $\hat{3}$ might simply be a manifestation of unstable-to-stable movement, and that the semitone relationship between the pitch pairs is merely coincidental.

Fortunately, the dual-scale system of major and minor modes offers an opportunity to control for this confound. Although the scale-tone hierarchies are regarded as similar between the major and minor scales, the positions of the semitones differ between the two scales. Specifically, in the major scale, semitone relationships exist between $\hat{3}$ and $\hat{4}$ and between $\hat{7}$ and $\hat{1}$. In the harmonic minor scale, semitone relationships exist between $\hat{2}$ and $\hat{3}$, between $\hat{5}$ and $\hat{6}$, and between $\hat{7}$ and $\hat{1}$. The contrasting placement of semitones in these two modes allows us, at least to some extent, to be able to examine semitone pitch movements relatively independently of the hierarchical importance of the different scale tones. That is, the contrast between the major and minor scales affords the opportunity to test our conjecture independent of the effect of scale degree.

II. HYPOTHESIS

In light of this background, we might propose the following specific hypothesis:

H. There is an association favoring semitone movement so that movement between $\hat{3}$ and $\hat{4}$ is favored in the major mode over the minor mode, while movement between $\hat{2}$ and $\hat{3}$ is favored in the minor mode over the major mode.

III. METHOD

In general, our method involves calculating the frequency of successions for various scale tones in a sample of major and minor-mode works.

A. Sample

Since our hypotheses relate to tone successions, an important sampling criterion is to focus on musical materials for which the linear succession of tones is not contentious. That is, we need to ensure that there is no ambiguity or dispute that tone X is followed by tone Y. Of the various musical textures, the least contentious would be musical melodies. Accordingly, in selecting our musical materials, we aimed to sample unambiguous musical melodies or thematic material.

For the purposes of this study, we employed two convenience samples. Specifically, we made use of two existing monophonic musical databases:

1. A random sample of 7,704 major and 768 minor-mode songs from the Essen Folksong Collection (Schaffrath & Huron, 1995).

2. A random sample of 7,171 major- and 2,618 minor-mode themes from the Barlow and Morgenstern *Dictionary of Musical Themes* (1948).

In both of these databases, the determination of the mode for each musical passage was made by the database authors. We have no information about the provenance or method by which these determinations were made. For the purposes of this study, we simply accepted the major and minor designations as encoded by the database authors.

Although a musical work might be nominally "in the major mode" or "in the minor mode," it is common for works to exhibit various deviant passages. In the minor mode, for example, it is common to encounter so-called "modal mixture" in which the major and minor modes co-mingle.

In addition, chromatic alterations are common in both major- and minor-mode passages. These modifications might introduce unanticipated confounds that could skew the results in various ways. It would be appropriate, therefore, to establish criteria by which certain musical works might be excluded from the sample.

Of particular concern would be those alterations that render a nominally major-mode work to more closely resemble the minor mode, or a nominally minor-mode work to more closely resemble the major mode. For example, any nominally major-mode melody that contains ♭3, or any nominally minor-mode melody that contains ♮6 would be suspect.

The main differences between the major and (harmonic) minor modes are found in scale degrees $\hat{3}$ and $\hat{6}$. Scale tone $\hat{7}$ is more problematic. In the minor mode, both ♭7 and ♮7 regularly appear and so it may be inappropriate to exclude any nominally minor-mode melody or theme either because it employs ♭7 or because it employs ♮7.

As a result, we resolved to exclude any nominally major-mode melody or theme that exhibits either ♭3, ♭6 or ♭7, and to exclude any nominally minor-mode melody or theme that exhibits either ♮3 and ♮6. Employing this criterion, 389 of the original 7,704 major-mode melodies and 205 of the original 768 minor-mode melodies were excluded from the Essen Folksong collection. Similarly, 1,547 of the original 7,171 major-mode themes and 704 of the original 2,618 minor-mode themes were excluded from the Barlow and Morgenstern collection. Hence, our final sample included 7,315 major- and 563 minor-mode songs from the Essen Folksong Collection Procedure, and 5,624 major- and 1,914 minor-mode themes from the Barlow and Morgenstern *Dictionary of Musical Themes.*

All of the sampled materials are available in the Humdrum "kern" format. The data were processed using the Humdrum Toolkit (Huron, 1994). Specifically, each melody was translated to a scale-degree representation, and then all of the scale-degree transitions were tallied. Since rests often indicate grouping boundaries, the relationship between pitches separated by a rest appears to be perceptually less salient. Accordingly, scale-degree transitions spanning a rest were omitted. In the Essen Folksong collection, phrases are explicitly notated. For the same reason, we omitted transitions occurring at phrase boundaries for this sample. That is, we did not consider the last note of one phrase to be "connected" to the first note of the ensuing phrase.

IV. RESULTS

Recall that our hypothesis predicts an association favoring semitone movement so that movement between $\hat{3}$ and $\hat{4}$ is favored in the major mode over the minor mode, while movement between $\hat{2}$ and $\hat{3}$ is favored in the minor mode over the major mode. Tables 1a and 1b present the pertinent tallies. Both tables show the total number of instances of movement between $\hat{2}$ and $\hat{3}$ and between $\hat{3}$ and $\hat{4}$. Table 1a pertains to the Essen Folksong collection; Table 1b pertains to the Barlow and Morgenstern themes.

Table 1a. Comparison of frequency of movement between $\hat{2}$ and $\hat{3}$ and $\hat{3}$ and $\hat{4}$ in the Essen Folksong collection.

	Major	Minor
$\hat{2} \leftrightarrow \hat{3}$	39,637	2,815
$\hat{3} \leftrightarrow \hat{4}$	19,685	2,515

Table 1b. Comparison of frequency of movement between $\hat{2}$ and $\hat{3}$ and $\hat{3}$ and $\hat{4}$ in the Barlow and Morgenstern themes.

	Major	Minor
$\hat{2} \leftrightarrow \hat{3}$	9,345	3,643
$\hat{3} \leftrightarrow \hat{4}$	8,014	2,544

The hypothesized association would predict that major/$\hat{3}\leftrightarrow\hat{4}$ and minor/$\hat{2}\leftrightarrow\hat{3}$ would exhibit higher tallies than major/$\hat{2}\leftrightarrow\hat{3}$ and minor/$\hat{3}\leftrightarrow\hat{4}$. An appropriate statistical test for this association is the chi-square test for contingency tables. In the case of the Essen Folksong Collection the results are not consistent with the hypothesis. In fact, there is a significant reverse relationship, $\chi^2(1) = 424.67$, $p < .01$; Phi = .08, Yates' continuity correction applied. Similar reverse results are evident in the Barlow and Morgenstern themes, $\chi^2(1) = 46.79$, $p < .01$, Phi = .05. In both cases the effect size is very small however.

V. DISCUSSION

We predicted an association favoring semitone movement so that movement between $\hat{3}$ and $\hat{4}$ in the major mode would be more common than $\hat{3}$ and $\hat{4}$ in the minor mode, while movement between $\hat{2}$ and $\hat{3}$ would be favored in the minor mode over the major mode. However, our results showed a significant (though very small) reverse association. Instead, activity between $\hat{2}$ and $\hat{3}$ tends to always be greater than activity between $\hat{3}$ and $\hat{4}$. In light of these results, the yearning theory appears to be weak.

In the course of this study, our analyses produced an unexpected result regarding the use of semitone intervals in melodies generally. In both the Barlow & Morgenstern and the Essen Folksong Collection, the most commonly sought-out intervals are the conjunct intervals of major and minor seconds. Descending seconds appear to be more sought-out than ascending seconds, and minor seconds are sought-out more than major seconds. Unisons also occur more frequently than a chance level, although to a lesser extent. A detailed comparison of the use of unisons, minor seconds, and major seconds in actual and scrambled melodies is presented in Tables 3a-d.

Table 3a. Comparison of semitone and whole-tone frequencies in scrambled and unscrambled melodies: Essen Folksong Collection, major mode.

Interval	Actual (%)	Scrambled (%)	Difference	Increase (%)
Unison	21.8	16.8	+ 5.0	29.8
Ascending minor second	6.2	3.3	+ 2.9	87.9
Descending minor second	8.3	3.3	+ 5.0	151.5
Ascending major second	13.3	10.1	+ 3.2	31.7
Descending major second	21.5	10.1	+ 11.4	112.9

Table 3b. Comparison of semitone and whole-tone frequencies in scrambled and unscrambled melodies: Essen Folksong Collection, minor mode.

Interval	Actual (%)	Scrambled (%)	Difference	Increase (%)
Unison	18.5	16.4	+ 2.1	12.8
Ascending minor second	9.1	4.3	+ 4.8	111.6
Descending minor second	12.4	4.3	+ 8.1	188.4
Ascending major second	16.4	9.9	+ 6.5	65.7
Descending major second	21.9	10.0	+ 11.9	119.0

Table 3c. Comparison of semitone and whole-tone frequencies in scrambled and unscrambled themes: Barlow and Morgenstern, major mode.

Interval	Actual (%)	Scrambled (%)	Difference	Increase (%)
Unison	14.7	13.9	+ 0.8	5.8
Ascending minor second	9.0	3.7	+ 5.3	143.2
Descending minor second	9.9	3.9	+ 6.0	153.8
Ascending major second	15.0	8.4	+ 6.6	78.6
Descending major second	18.3	8.5	+ 9.8	115.3

Table 3d. Comparison of semitone and whole-tone frequencies in scrambled and unscrambled themes: Barlow and Morgenstern, minor mode.

Interval	Actual (%)	Scrambled (%)	Difference	Increase (%)
Unison	14.7	13.6	+ 1.1	8.1
Ascending minor second	11.2	5.0	+ 6.2	124.0
Descending minor second	13.4	5.0	+ 8.4	168.0
Ascending major second	13.0	7.3	+ 5.7	78.1
Descending major second	16.1	7.2	+ 8.9	123.6

In general, there are more whole-tone steps than semitone steps in melodies. However, compared with scrambled reorderings of tones, in actual melodies semitone intervals are favored more than whole-tone intervals. In all four samples (Tables 3a-3d) the percentage increase for ascending and descending semitones is greater than the corresponding increase for whole-tone intervals. This implies that close pitch proximity is favored.

VI. CONCLUSION

We investigated some 15,000 melodies and themes from the Essen Folksong Collection and the Barlow and Morgenstern *Dictionary of Musical Themes* in order to test whether there is an association favoring movements between scale degrees $\hat{3}$ and $\hat{4}$ (semitone) over $\hat{3}$ to $\hat{2}$ (whole-tone) in the major mode, compared with movements between scale degrees $\hat{3}$ and $\hat{4}$ (whole-tone) over $\hat{3}$ to $\hat{2}$ (semitone) intervals in the minor mode. The results were not consistent with the “yearning semitone” theory and failed to show the predicted relationships.

However, *post-hoc* observations show that while whole-tone intervals outnumber semitone intervals, composers nevertheless exhibit an even stronger affinity for using semitone intervals in general. By way of summary, our results are not consistent with the “yearning semitone” theory, but our study does offer *post-hoc* evidence consistent with an alternative theory—what might be called the “baby steps” theory: the smallest pitch movements appear to be favored whether or not these movements are linked to tonally more stable pitches. These results reinforce longstanding observations regarding pitch-proximity, but add to these observations by identifying a disposition towards using the smallest available intervals in the construction of melodies.

Of course these observations may not generalize beyond the specific repertoires studied. Further study is warranted to establish whether “baby steps” are preferred in other styles of Western melody, and whether the theory might apply to melodies from non-Western cultures.

REFERENCES

Arthur, A. (2015, October). A comprehensive investigation of scale-degree qualia: a theoretical, cognitive and philosophical approach. Paper presented at the 38th Annual Meeting of the Society for Music Theory, St. Louis, MO.

Barlow, H., & Morgenstern, S. (1983). A dictionary of musical themes. New York: Crown Publishers. Revised ed. London: Faber & Faber. (Original work published 1948)

Bharucha, J. J. (1984). Anchoring effects in music: The resolution of dissonance. *Cognitive Psychology*, *16*(4), 485-518.

Bharucha, J. J. (1996). Melodic anchoring. *Music Perception: An Interdisciplinary Journal*, *13*(3), 383-400.

Cooke, D. (1959). *The language of music*. London: Oxford University Press.

Deutsch, D. (1978). Delayed pitch comparisons and the principle of proximity. *Perception & Psychophysics*, *23*(3), 227-230.

Fétis, F. J. (1853). *Traité complet de la théorie et de la pratique de l'harmonie contenant la doctrine de la science et de l'art.* Brandus et Cía.

Francès, R. (1988). *The perception of music*. (Dowling, J., Trans.). Psychology Press. (Original work published 1958)

Fuller, S. (2011). Concerning Gendered Discourse in Medieval Music Theory: Was the Semitone “Gendered Feminine?”. *Music Theory Spectrum*, *33*(1), 65-89.

Huron, D. (1994). The Humdrum Toolkit: Reference manual. Menlo Park, CA: Center for Computer Assisted Research in the Humanities.

Huron, D. (2001). Tone and voice: A derivation of the rules of voice-leading from perceptual principles. *Music Perception: An Interdisciplinary Journal*, *19*(1), 1-64.

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. MIT press.

Krumhansl, C. L. (1990). Tonal hierarchies and rare intervals in music cognition. *Music Perception: An Interdisciplinary Journal*, *7*(3), 309-324.

Krumhansl, C.L., & Kessler, E.J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial map of musical keys. Psychological Review, 89, 334-368.

Krumhansl, C. L., & Shepard, R. N. (1979). Quantification of the hierarchy of tonal functions within a diatonic context. *Journal of experimental psychology: Human Perception and Performance*, *5*(4), 579-594.

Leach, E. E. (2006). Gendering the Semitone, Sexing the Leading Tone: Fourteenth-Century Music Theory and the Directed Progression. *Music Theory Spectrum*, *28*(1), 1-21.
Lerdahl, F. (2001). *Tonal pitch space*. Oxford University Press.
Lussy, M. (1892). *Traité de l'expression musicale*. Libraire Fischbacher.
Schaffrath, H., & Huron, D. (1995). The Essen folksong collection in the humdrum kern format. *Menlo Park, CA: Center for Computer Assisted Research in the Humanities*.
Yeary, M. J. (2011). Perception, pitch, and musical chords. (Doctoral dissertation). Retrieved from ProQuest Dissertations Publishing (Order No. 3472979)

Statistical learning of novel musical material: Evidence from an experiment using a modified probe tone profile paradigm and a discrimination task

A.-X. Cui[1*], C. Diercks[1], N.F. Troje[1], L.L. Cuddy[1]
[1]Queen's University, Kingston, ON, Canada
[*]corresponding author

Background & Aims
Statistical learning is thought to play a crucial role in the process of gaining musical knowledge. Our aim is to explore this role. We asked: are participants able to abstract statistical (pitch distributional) information from novel melodies after 30 minutes of exposure?, does this abstraction become a mental representation?, and does this representation influence responses to melodies generated from a similar, but not identical, pitch distribution?.

Method
We assessed statistical learning before and after exposure using the probe tone method. Participants indicated the goodness of fit (GoF) of each of 12 probe tones to novel melodies (contexts). Melodies heard during exposure and as contexts were both based on a whole-tone scale, but differed in 12% of the presented tones. This allowed us to investigate GoF responses to probe tones that occurred both during exposure and in contexts (EyPy), probe tones that occurred in the contexts, but not during exposure (EnPy), probe tones that occurred during exposure, but not in contexts (EyPn), and probe tones that occurred neither during exposure or in contexts (EnPn). Then, participants completed a discrimination task, in which they indicated which of two melodies resembled the heard melodies.

Results & Conclusions
As expected, prior to exposure, GoF responses were higher for probe tones appearing in the novel melodies than probe tones not appearing in the novel melodies. This finding also held post exposure, but in addition, GoF responses for probe tone category EyPn increased significantly after exposure. This indicates that participants were able to abstract pitch distributional information during 30 min of exposure as well as when readily accessible (heard in the probe tone context), and integrate this information when making GoF responses. The average percent correct of 76% in the discrimination task shows that participants were able to transfer the gained musical knowledge to another task.

Robust training effects in congenital amusia

Kelly L. Whiteford[1] and Andrew J. Oxenham[1,2]
[1]Department of Psychology, University of Minnesota, Minneapolis, MN, USA
[2]Department of Otolaryngology, University of Minnesota, Minneapolis, MN, USA

Congenital amusia, commonly known as “tone deafness,” is a neurogenetic disorder in music perception, related to an inability to discriminate fine-grained changes in pitch. The general consensus is that amusia is a life-long deficit and does not improve with training; however, few studies have directly tested this assumption. Our aim was to determine whether pure-tone pitch perception in amusia can improve with practice and, if so, whether the improvement generalizes to 1) other non-trained pure-tone frequencies and 2) music perception, as measured with the Montreal Battery of Evaluation of Amusia (MBEA). Forty subjects (20 amusics) participated in a 4-day training paradigm where they trained on either 500-Hz pure-tone pitch discrimination or a control task involving the lateralization of broadband white noise based on interaural level differences (ILDs). Pure-tone pitch perception at 500 Hz, 2000 Hz, and 8000 Hz and the MBEA were was assessed pre and post training. Results were consistent with previous studies, showing that amusics had impaired pitch perception but normal sound localization. All participants improved with training on their respective tasks. The bulk of this training occurred between sessions 1 and 2 for both pitch and localization tasks. Analyses of pre- and post-training data indicated improved pitch perception across all frequencies for all groups, even those who trained on the non-pitch control task. Surprisingly, 11 out of 20 amusics no longer met the diagnostic criterion for amusia based on the MBEA after the training. Our preliminary results reveal robust pitch training effects in both normal controls and amusic subjects that generalize across frequencies and to more complex tasks, involving melody discrimination. This suggests at least some plasticity in the perception of pitch in amusics, leading to some previously amusic subjects to no longer meet the currently accepted criteria for amusia following training.

Listeners' perception of rock chords: Further evidence for a more inclusive harmonic hierarchy

Lincoln Craton[1], Peter Krahe[1], Gina Micucci[1] and Brittany Burkins[1]
[1]Stonehill College, Easton, MA, USA

Rock harmony diverges from the conventions of common-practice music. Specifically, although both common-practice music and a lot of rock music draw from the "basic diatonic set"—that is, triads built on the diatonic scale—rock routinely employs many other chords as well (de Clercq & Temperley, 2011; Stephenson, 2002). We recently reported preliminary evidence that listeners prefer some chords that are common in rock, but that lie outside the basic diatonic set, to atypical chords; this effect was strongest for major chords (Craton, Juergens, Michalak, & Poirier, in press, *Music Perception*). The present research replicates and extends these findings.

We adapted a rating task developed by Krumhansl (1990) in her pioneering studies of the common-practice harmonic hierarchy. Participants were recruited via Amazon's Mechanical Turk (MTurk) and directed to an online experiment hosted by Qualtrics. In Experiment 1 (N = 199), participants provided liking ratings (1 = *dislike extremely*; 10 = *like extremely*) for major chords that followed a short key-establishing passage (a major scale + tonic major triad in root position). Twenty target chords were presented in root position and included every chromatic root from IV (descending) to octave (ascending). The effect of chord on liking ratings was significant, $F(14.32, 2835.52) = 45.13$, $p < .001$, partial $\eta^2 = 0.186$. Pairwise comparisons revealed a pattern of mean liking ratings that closely replicate our earlier findings.

Experiment 2 (N = 188) was identical, except that participants rated how well each target "fit" with the preceding musical context. The effect of chord on fitness ratings was significant, $F(15.39, 2878.53) = 73.47$, $p < .001$, partial $\eta^2 = 0.282$. Pairwise comparisons revealed a similar pattern, except that fitness ratings for the tonic were particularly high. These finding provide further support for the claim that listeners experience rock chords as musically appropriate.

Comparative judgement of two successive pitch intervals

Makiko Sadakata,[1,2,3] Maartje Koning,[3] Kengo Ohgushi[4]

[1] *Institute for Logic, Language and Computation, University of Amsterdam, The Netherlands*
[2] *Music department, University of Amsterdam, The Netherlands*
[3] *Artificial Intelligence department, Radboud University Nijmegen, The Netherlands*
[4] *Kyoto City University of Arts, Japan*

[1] m.sadakata@uva.nl, [2] koningmaartje@gmail.com, [3] ohgushi.k@view.ocn.ne.jp

I. INTRODUCTION

Variable acoustic cues are available to listeners as music unfolds over time. Among these cues, listeners greatly rely on relative pitch intervals that are presented either simultaneously or sequentially in music, which form foundations of our perception of melody and harmony. While the importance of such relative pitch processing in everyday music listening has repeatedly been emphasized, thus listeners are suggested to be more sensitive in perceiving relative pitch information than e.g., absolute pitch processing (Honing 2011), it was demonstrated that listeners find it rather hard to compare interval sizes of two successive tone pairs (Sadakata & Ohgushi, 2000). The current study elaborates on this finding by testing how tonal contexts and listeners' musical experience interact with the task performance.

II. Methods

Twenty-seven participants took part in 3 tests: the comparative judgement of pitch intervals (comparison experiment), part of the Goldsmith-MSI (excluding the listening part, Müllensiefen et al., 2014) and the pitch discrimination threshold test.

The stimuli of the comparison experiment consisted of two pure tone pairs (t1-t2 and t3-t4). Participants compared whether the size of pitch interval of t3-t4 was smaller, the same or larger than that of t1-t2.

The stimuli were designed to compare the following 4 factors: 1. within pair pitch contour (e.g., up or down within t1-t2 and t3-t4), 2. between pair pitch contour (600 cents up or down from t1 to t3), 3. size of pitch intervals (small/medium/large each for t1-t2 and t3-t4) and 4. interval familiarity (common/uncommon). Common interval sizes were 300, 700 and 1000 cents, and uncommon intervals were 250, 650, 950 cents, respectively. The included pitch ranged from 184 Hz to 987 Hz.

The 75% pitch discrimination threshold from C4 was determined using the forced choice discrimination task with the staircase procedure.

III. Results

The overall correct response rate of the comparison experiment was about 40%, confirming the previous finding that the task is challenging. Further analyses revealed that, when pitch contours within and between pairs were in the same direction (e.g., pitch contour from t1-t2, t3-t4 and t1-t3 were all upwards), listeners tended to overestimate the size of t3-t4 relative to t1-t2. Conversely, in general, when directions of between and within pairs disagreed, listeners tended to underestimate the size of t3-t4 relative to t1-t2. This trend tended to be exaggerated in larger intervals. The interval familiarity did not result in any significant effects. Interestingly, individuals with higher Goldsmith-MSI and pitch discrimination performance tended to perform better in this task, but only when intervals were small and the t1-t3 pitch contour was ascending.

IV. Discussion

Although processing of relative pitch intervals is one of the most important skills in our daily musical experience, the current study demonstrated that comparing size of two pitch intervals is challenging. This may be because global pitch contours interfere with judgements of local pitch interval sizes. Interestingly, the same results patterns were observed for pitch intervals that are both common and uncommon to the listeners, suggesting that this interference effect is not specific to Western tonal music context.

REFERENCES

Honing, H. (2011). Musical Cognition. A Science of Listening. New Brunswick, N.J.: Transaction Publishers.

Müllensiefen, D., Gingras, B., Stewart, L. & Musil, J. (2014). The Musicality of Non-Musicians: An Index for Measuring Musical Sophistication in the General Population. *PLoS ONE* 9(2): e89642.

Sadakata, M & Ohgushi, K. (2000). Comparative judgements pitch intervals and an illusion. [Ontei no hikakuhandan to sore ni tomonau sakkaku genshou]. *Proceedings of the Annual meeting of the Japanese Society for Music Perception and Cognition*, pp. 35-42, Sendai.

The influence of diatonicity, tone order, and composition context on perception of similarity: An experiment and replication

Mark Shevy[1], Douglas McLeod[2], Stephen Dembski[3]

[1]Communication and Performance Studies, Northern Michigan University. Marquette, Michigan, USA

[2]School of Journalism, University of Wisconsin, Madison, Wisconsin, USA

[3]School of Music, University of Wisconsin, Madison, Wisconsin, USA

There is a need for more nondiatonic music perception research (cf. Krumhansl & Cuddy, 2010). In an experiment (N=177) and a replication (N=105), listeners were randomly assigned to hear a series of pairs of music. Following each pair, they rated the similarity of the two pieces. The independent variables of the study consisted of three musical manipulations: 1. Compositional context is based upon the original composition used for creating musical stimuli variations. One set of variations was created by systematically altering the nondiatonic piece *Alta* (Dembski, 2001). The other set of variations was created by altering *Well-Tempered Clavier Prelude in C major* (*WTC*). 2. The second manipulation was diatonicity (diatonic pitch set vs. nondiatonic). 3. The third manipulation was tone order (*Alta* tone order vs. *WTC* tone order).

The results indicate that for all pieces of music, diatonic and nondiatonic, listeners rated the pieces in an identical pair (the same piece of music presented twice) as being much more similar than pieces in a pair that differed in diatonicity and/or tone order. This finding adds to the existing empirical evidence of what some scholars assume is self-evident, yet other people still question – that listeners can perceive structure in nondiatonic music. The results also indicate that changes in diatonicity resulted in significantly lower similarity ratings than changes in tone order, if the music was in the *Alta* composition context. Conversely, in the *WTC* composition context, changes in tone order caused the greater decrease in perceived similarity. This pattern of results was strikingly consistent in the replication experiment. The results will be discussed in terms of tonal cognitive schemas, probabilistic learning, and the possibility that the rules established by a composer may, in certain contexts, have more bearing on perception than the mental models developed through general music exposure.

The Alberti Figure Visualized Challenges the Stability of Tonicity

Nancy Garniez

***Abstract*—**
Since first reading *Sound and Symbol* by Viktor Zuckerkandl in 1963 I have built a life of performance and teaching upon his observations about the logic of intuitive hearing. My goal is to explore the extent to which acknowledged variable responses to tone and tone relatedness lead to confident, engaged performance. I outline the chronology and the methodology of my work, starting with the process of freeing myself from the constraints imposed by standard music notation upon my trained ear. Learning to hear as amateurs and children hear has enabled me to lead them to value the ways in which their intuitive responses--"wrong" notes and "faulty" intonation, *e.g.*,--may reveal a composer's process. Three distinct areas of pitch perception are involved: *a cappella* singing, chamber music, and solo piano--by far the most difficult to observe. I briefly describe Tonal Refraction, the visualization technique to which the work gave rise in 1993, detailing how visualizing Mozart's *Piano Sonata in G*, K. 283 was pivotal in connecting a vivid memory of child hearing to my mature artistic ear. The Alberti figure, which had previously seemed uninspired, proved seminal in comprehending the difference between habitual responses based on learned notions of tonicity, and full connection with the piano's infinitely variable complex sound. One might study how alternative visualizations assist in accounting for the long as well as short-term processing of tone. Another topic is the study of subjective reactions to specific pitches in specific compositions. Also of interest is the way in which color selection reveals an individual's sense of tone relatedness.

***Keywords*—**Alberti figure, Mozart, Visualization, Zuckerkandl

INTRODUCTION

The experiment that culminated in this finding began some 50 years ago as a result of reading Viktor Zuckerkandl's *Sound and Symbol*. in which he demonstrates that the ear, whether trained or not, makes sense of tones. "The acoustically wrong tone can be musically right if the deviation is right in the sense of the movement. It cannot, then, be simply tones as such, tones of a predetermined pitch, which we sing or hear in melodies:it is motion represented in tones" [1].

I had been aware since first encountering remarkable sounds in Mozart sonatas as a young child that there were qualities intrinsic to my hearing which remained unaddressed throughout my musical training, qualities made memorable by what I now know to be pre-conscious responses. Sensing that Zuckerkandl's work related directly to those qualities, I resolved through teaching to prove its validity. Pursuing that deeply personal fascination with listening entailed identifying responses within three distinct musical contexts with which I had been involved as a child: the sound of the piano, choral singing. and chamber music. Activating my awareness of these enabled me to objectify aspects of my own pre-conscious hearing in the form of a visualization method, Tonal Refraction.

The Alberti figure proved a powerful clarifying tool already at the method's earliest manifestation, as it directed attention to the ways in which the figure's several voices defined and energized distinct registers of the piano. Many years of exploring the method revealed the extent to which habitual hearing of theoretically identifiable harmonies was inhibiting my responses to the overtone activity generated by the Alberti figure. Now, no longer a tedious, predictable cliché, it becomes central to a lively, endlessly variable approach to the Mozart piano sonatas, one of which provoked this curiosity many years ago.

CHRONOLOGY

That the sound of the piano was special to me was clear from my first experience of the instrument as a toddler. I had never seen or heard a piano other than the spinet in a neighbor's apartment, to which I was given free access. At twelve my deep love of the sound met its first match via the printed page in certain sounds of the Mozart *Sonata in G*, K. 283. But my involvement with piano sound was interrupted twice: first, at fifteen, when I quit piano for reasons to this day unclear and organ became my default instrumental focus through college and a Fulbright year; and again at twenty-three when, despite three years of intensive piano study, a congenital spinal deformity seemed to preclude professional solo performance. It was at this point that Carl Schachter suggested I read *Sound and Symbol* by Viktor Zuckerkandl and the ongoing experiment began.

METHODOLOGY

The goal throughout is to hear as students hear, not as I think they should hear. I did this in relation to the three contexts of listening: *a*) *a cappella* singing; *b*) chamber music for diverse instruments; and *c*) solo piano. The students whose listening I observed were well suited to the experiment, as they were either private students, young people at the Mannes College of Music Preparatory Division, or adult amateurs, primarily participants in the Chamber Music program I coordinated at the Mannes Extension Division

These populations, being less practiced and less pressured, gave uninhibited expression to their spontaneous aural responses. I wanted to take their deviations seriously enough to guide them into the composer's logic while building confidence in their natural auditory inclinations.

a) Starting with the sound in which the listening ear is most clearly manifest, in 1975 I undertook an experiment to teach adult amateurs to sight-sing in tune. A warm-up exercise focussing on the difference between random sounds and a tunable pitch assured the group's visceral awareness of vocal intonation. Based on Zuckerkandl's observations about the major scale in relation to the overtones of the tonic [2]. I used the staff to locate the 1st, 3rd, and 5th degrees of the scale of each piece, noting that only these pitches need be in tune, leaving the others indeterminate. Thus "tuning" the staff connected the act of reading directly to awareness of linear *vs*. harmonic tuning. The experiment continued with many of the original participants for over twenty-five years; the foreshortened staff became the basis of the vertical axis of my visualization method.

b) Objective listening to ensemble playing proved remarkably difficult, as it required that I listen without recourse to a printed score in order to perceive lapses in intonation and wrong notes not as errors, but as deviations from the composer's version of a sensible line. The process had an unforeseen result: not only was it evident that certain notes or passages provoked deviations, but that other specific notes or passages were remarkably accurate and in tune. It became apparent that much of a work's inherent drama stemmed from the composer's grasp of the acoustical specificity of the instruments.

The possibility that instrumental acoustics played a structural role in Classical composition seemed at once self-evident and arcane, as the identifying quality of those passages became evident not from analysis of the printed score or by conscious intention of the players, but from listening to uninhibited playing. (This was subsequently borne out in contrasting Mozart's writing for piano in the sonatas for piano and violin, K.301-306, with those for piano solo, K. 279 -285.

Nancy Garniez, Mannes School of Music, New School University New York, NY 10014 phone: 212-749-4035; e-mail: garniezn@newschool.edu

c) Of the three listening "modes" the solo piano proved the most difficult to observe. Visualizing Mozart sonatas over a period of many years leads me to conclude that this is because of the clash between the complexity of actual piano sound with the readily analyzable harmonies of standard music notation.

In 1993, when an innovation in Physical Therapy enabled me to resume solo performance, decades of immersion in ensemble playing had heightened my awareness of acoustical properties distinct to the solo piano. This awareness manifested itself in two areas:

1) The complexities of equal-temperament so consumed my ear that I could no longer find the relatively pure overtones of mean-tone intonation with which I had been experimenting on the harpsichord.

2) Concentration on touch and fingering revealed with clarity the difference between the resonance of white and black keys, the result of the piano's unique levered action, whereby the amount of audible overtones in a tone is directly affected by the position of the fingers on black keys relative to white.

Celebrating my return to the solo piano, I programmed a recital on which I played the Mozart *Sonata in G*, K. 283, the work in which specific sounds had caught my attention as a young child. I found myself awake in the middle of the night convinced that now I could *show* what it was about this work that had so fascinated me all these years. Using the crude colored pencils I happened to have on hand and graph paper I visualized the way I heard the piece [Fig. 1].

Though here the vertical axis of the grid is quite primitive, two features stand out: 1) the Alberti figure is clearly identifiable; 2) the colors for B and D are much more vivid than that for G. This entirely unconscious selection of colors surprises me after many years, as it represents a sense of the relative values of the three tones that can have been informed only by my memory of what I had heard as a child. Notable also is the solidity of the first tone, D (blue), and the brightness of B (red), in contrast to the almost invisible G (light gray). Coming into a composition via its openings tones rather than in the context of as-yet-unheard tonicity has proved one of the insightful strengths of Tonal Refraction.

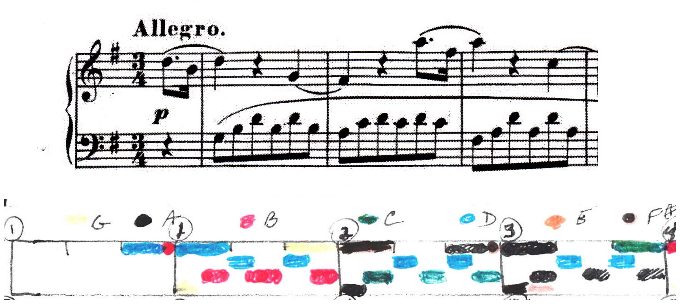


Fig. 1. Very first visualization, 1993: Mozart K. 283, mm1-3.

I was sufficiently troubled by the conflict between what I thought I should hear and what I had actually heard before formal theory training that it was a long time before I could discern the real content of this first visualization. The values conforming to clear tonicity, *i.e.*, how I had been trained to hear, would better be depicted in the two blocks of contrasting color, green for the tonic, red for the dominant [Fig. 2].

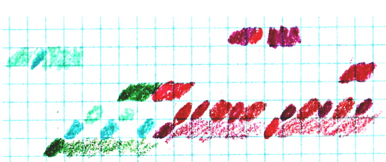


Fig. 2. The passage as heard harmonically with distorted bass values.

Influenced by both the visual clarity and the auditory sense of the Alberti figure, I began basing the vertical axis of each grid on the "tuned" staff of the sight-singing class, *i.e.*, the 1st, 3rd, and 5th degrees of the tonic, with intervening pitches placed subjectively as they felt closer to one or the other of these orientation pitches, or "stations" of the octave. The pitches of the composition were represented with colors, chosen according to whatever plan seemed to fit my hearing of the piece that day. The horizontal axis represented metric values, the rigid lines of the grid heightening awareness of articulations.

The Alberti figure provided the single most informative visual element that revealed the extent to which my hearing of Mozart sonatas had become pre-conditioned by learned notions of tonicity. Of the nineteen Mozart solo piano sonatas, eleven use the Alberti figure within the first eight bars; in only two of these does the tonic occur prominently as the repeated tone, probably to demonstrate the metric vitality of the off-beat repeated tone of the figure.

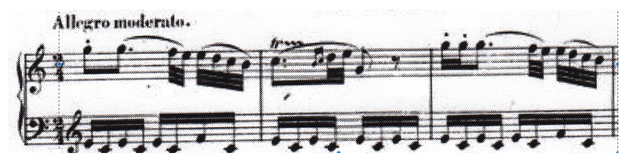


Fig. 3. K. 330 with reverse Alberti figure, emphasizing tonic.

Simply by noticing the visual effect of the constantly repeated off-beat tones of the Alberti figure I became aware of the rhythmic variation from one measure to the next caused by the reference of overtones in the Alberti figure to the usually longer melody tones, the most obvious of which are indicated by diagonal lines in Fig. 4. Note that these melody tones are longer than the bass notes.

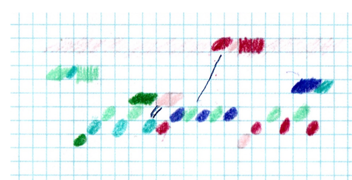


Fig. 4. Diagonal lines indicate the obvious overtones relating Alberti tones to melody.. The F# overtone hovers above the "real"notes.

By focussing on such clear organizing factors as root tones and downbeats, I had been ignoring the myriad variables of the piano's sound that had intrigued me as a child, giving me the distinct feeling that the entire G major Sonata was about the contrast between the black key F# and the white key F-natural. Now, seeking out these variables in large measure via the Alberti figure, I find new life in the sonata for piano solo. Repetition of a piece or section is acoustically impossible.

It was only after several years and the involvement of many individuals in the visualization process that its full sense and implications became clear [3]. Perhaps its greatest value is its power to objectify subjective elements of music perception, removing all fear of right or wrong.

CONCLUSION

From the earliest stages of music learning visual imagery and expression have a profound impact upon how we hear and on our understanding of what we hear [4]. Musical tone, by nature complex, cannot be depicted with discrete symbols on a staff. Even young children seem reassured when they learn of the 32,000 vibrating hair cells in the inner ear [5]. A visualization system with room for subjectivity allows musicians of every age and level to accept and explore various dispositions of tone and of rhythm as processed holistically. in ways that may or may not conform to prior experience or to learned concepts [6].

Visualizing the Alberti figure reveals its power to lift the ear into the realm of overtones, thereby enabling us to relish the Mozart solo sonatas as celebrations of the piano's unique acoustical riches.

IMPLICATIONS FOR FUTURE STUDY

One might study subjective factors in response to specific pitches in compositions; the effect of using an alternative notation system to account for the long as well as short term processing of tone; the way in which color selection reveals an individual's sense of tone relatedness..

ACKNOWLEDGMENT

Thanks to the many individual donors and colleagues who have supported this research, its publication, and its presentation in conferences; to past and present students, including Dr. Pamela Stanley, (The Horace W. Goldsmith Chair, Albert Einstein College of Medicine), Dr. Marlon Feld (Music Theory and Composition, Columbia University), Dr. Damon Horowitz (Philosophy, Columbia University), Thanks to Dr. Joan Farber, Clinical Psychologist and painter, for reading the paper; to Jacob Garniez for editorial and technical assistance.

Thanks to Mannes College of Music for a Faculty Development Grant; to Dr. George Fisher, former Dean of the Mannes College Division, for support and encouragement; and to Carl Schachter, without whom I would not have known of Viktor Zuckerkandl,

REFERENCES

1. Viktor Zuckerkandl, *Sound and Symbol*, *Vol. I. Music and the External World* (*tr.* Willard Trask), New York, NY, Pantheon Books, 1956, p. 81.
2. Zuckerkandl, *The Sense of Music* (corrected edition), Princeton, NJ, Princeton University Press, 1971, pp.25-27.
3. Garniez, Nancy, "Can Individuals, Both Trained and Untrained, Be Made Aware of Their Subconscious Responses to Tone and, If So, Would It Make Any Difference?"" *Book of Abstracts ICMPC 13-APSCOM 5*, 2014, p. 111.
4. Jeanne Bamberger, The Mind Behind the Musical Ear, Cambridge, MA, Harvard University Press, 1991, pp. 46, ff.
5. James Hudspeth,. "How the Ear's Works Work" in *Nature*, Vol. 341, 5 October, 1989.
6. Garniez, "Can One Teach Rhythmic Flexibility to Young Learners?" in *Clavier Companion*, Nov.-Dec, 2011, pp. 42-5.

Understanding modulations through harmonic priming

Olivia Podolak[1] and Mark A. Schmuckler[1]
[1]Department of Psychology, University of Toronto Scarborough, Toronto, Ontario, Canada

The harmonic hierarchy has been studied vigorously over the past few decades. However, few studies have scrutinized the effects of tonal modulations, or key changes, in regards to harmonic processing, despite the their relevancy in everyday listening. The present study used an adapted version of the classic harmonic priming paradigm, in which the length of the priming context was extended to include two keys connected by a single pivot chord. Fifteen musicians and seventeen non-musicians were asked to respond to a single target chord that was related to the first key, the second key, or neither key in the priming context. Two dependent variables were measured: (1) participants' reaction time to whether or not the target chord fit with the priming context, and (2) their subjective rating of the degree to which this target chord fit with the priming context. The results of the reaction time data showed that non-musicians were not able to distinguish between the varying degrees of relatedness shared by the prime and target. In contrast, musicians responded more quickly to targets that were related to the second key as opposed to targets related to the first key. Furthermore, musicians did not respond significantly faster to targets related to the first key versus those related to neither key. The results of the rating data showed that the highest ratings were given to targets that were related to the second key. Interestingly, however, targets related to the first key were still rated significantly higher than those related to neither key. Overall, the present study demonstrates the online flexibility of harmonic representations. The findings suggest that key changes operate as a function of recency rather than primacy, as participants rapidly disregarded the preceding key during modulation. Future research will be thus utilizing electroencephalography to untangle the sensory and cognitive factors that might be underlying these findings.

The use of combined cognitive strategies to solve more successfully musical dictation

Ruth Cruz de Menezes[1,2], Maria Teresa Moreno Sala[1,2], Matthieu J. Guitton[3,4]
[1]Laboratoire de recherche en formation auditive et didactique instrumentale (LaRFADI), Faculty of Music, Laval University, Quebec City, Canada
[2]Observatoire International de création et recherche en musique (OICRM), Canada
[3]Faculty of Medicine, Laval University, Quebec City, QC, Canada
[4]Institut Universitaire en Santé Mentale de Québec, Quebec City, Canada

Students often present difficulties in music melodic dictation, particularly in the resolution of some kind of intervals. For that reason, the purpose of this research was to analyze the type of strategies students used to solve intervals, which strategies led to succeed and which led to failure. To study that question, undergraduate students (n=48) were asked to perform a tonal melodic dictation during which they had to write down the cognitive strategies they were using throughout the process. Considering the example was a simple melody, we focused our analyses on the three intervals presenting more difficulties to students: a descending major third, an ascending perfect fourth and an ascending minor sixth. From the students' dictation descriptions, two types of main strategies emerged: non-tonal strategies (or primary) and tonal strategies (secondary). Afterwards, we analyzed each strategy to identify each interval, given a categorical binary result for each (successful/not-successful). Grouping together the results of the three intervals for all the participants (48), significant analyses ($\rho=0.52$ $p=0.0001$) were found between the number of strategies used and successful results in interval identification. The resolution of intervals failed only when no cognitive strategy was described (75% of failures). Conversely, when using one strategy 75% of subjects successfully solved the interval. Moreover, significant results were found between the type of strategies used and successful results in interval identification ($\rho=0.312$ $p=0.0001$). Subjects using one non-tonal strategy achieved a score of 64%, while those using a tonal strategy attained 100%. Finally, the combination of two or more tonal strategies always achieved 100%. To summarize, training students to combine different cognitive strategies to solve some intervals of a melody may be helpful for improving their skills in melodic dictation.

“Looseness” in musical scales: Interplay of style and instrument

Rytis Ambrazevičius

Department of Audiovisual Arts, Kaunas University of Technology, Lithuania
Department of Ethnomusicology, Lithuanian Academy of Music and Theatre, Lithuania
rytisamb@gmail.com

ABSTRACT

In the case of non-fixed or changeable intonation, musical scales demonstrated in practice show some acoustic “looseness,” roughly in two ways. First, there is some freedom of retuning for otherwise analogous performances resulting from JND and flexibility of pitch categories. Second, local inflections of intonations are common in form of systematic deviations (performance rules) and performance noise.

To analyze dependence of factors determining “looseness” in musical scales on musical style and instrument, scales of 76 vocal and *kanklės* (a type of psaltery) performances recorded in the 20th century and representing several styles of Lithuanian traditional music were measured and analyzed.

Different instances of interplay between the instrumental/vocal constraints and musical style were registered. *Kanklės* tunings used for newer diatonic melodies show relatively steady and similar patterns. The related sung melodies show more “loose” scales, which seems natural because of the differences in vocal / fixed instrumental intonation. However, genetically more archaic, appr. equidistant (~anhemitonic) scales in vocal and instrumental performances of *sutartinės* (a type of *Schwebungsdiaphonie*) show the reverse. The basic dyads in vocal *sutartinės* tend to be more steadily intoned, supposedly because of striving for maximum psychoacoustic roughness. This is less relevant in the instrumental case since the discussed playing techniques do not allow the realization of sustained rough dyads or clusters.

Hence “looseness” in musical scales depends both on the technically possible steadiness of intonation and on stylistic (incl. psychoacoustic) constraints. Phylogenetic aspects of this phenomenon could also be traced.

I. INTRODUCTION

When there is an option of retuning or free intonation in performance practice musical scales show some acoustic “looseness.” First, the “looseness” results at least from pitch JND which is only several cents under the best listening conditions (cf. Hess, 1983; Parncutt & Cohen, 1995, p. 836; Fastl & Zwicker, 2007, p. 186). In some cases pitch categories can gradually “migrate” resulting in so-called “unfolding scales” and similar instances (cf. Alexeyev, 1976, p. 48-58, and Ambrazevičius, Budrys, & Višnevska, 2015, p. 187-189, for traditional music); yet they are considered constant in the performers’ emic systems. Second, local inflections of intonations are common in the form of systematic deviations (performance rules) and performance noise (cf. Fyk, 1994; Kopiez, 2003; Friberg, Bresin, & Sundberg, 2006; Ambrazevičius & Wiśniewska, 2008).

In the present study, I limit myself mostly to the case of retuning flexibility as the result of flexible pitch categories, yet considered emically constant. My samples for the acoustic measurements come from Lithuanian traditional music. Therefore, I first present an extended introduction into aspects of Lithuanian traditional music relevant for the study.

A. Lithuanian Traditional Singing Styles

Even though Lithuania is a small country, it is traditionally divided into four main ethnographic regions. These divisions are based on spoken dialects, traditions, and other cultural elements: Aukštaitija (east), Žemaitija (west), Dzūkija (southeast), and Suvalkija (southwest) (Figure 1). Musical dialects approximately correspond to these ethnographic divisions.

Figure 1. Ethnogaphic regions of Lithuania

Individualized solo performance, monophony (heterophony), polyphony, and homophony are and were characteristic of various Lithuanian singing dialects. Monophony (heterophony) in ensemble performance is traditionally attributed almost exclusively to Dzūkija and Suvalkija. Nowadays, homophony is the prevailing style in different Lithuanian regions. Usually, one singer performs the leading part while the rest of a group adds the lower “background” part, thereby creating dyads of thirds, fourths, or fifths with the leading part, according to the functional harmony (Figure 2).

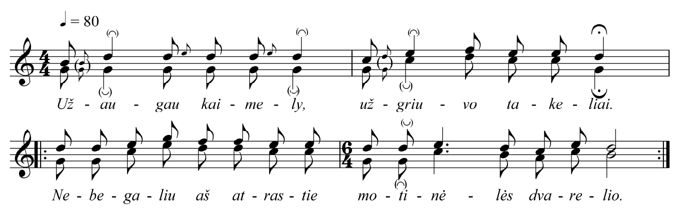


Figure 2. Schematic transcription of the song “Užaugau kaimely” (Jūrė, Marijampolė Dst., Suvalkija)

Sutartinės (sing. *sutartinė*) constitute an important part of Lithuanian polyphonic song lore. Many, but not all, *sutartinės* can be regarded as a kind of *Schwebungsdiaphonie* (beat diaphony). They are based on polyphonic and polyrhythmic patterns resulting from intertwining vocal parts. Typically, two voices perform simultaneously, creating sequences of chords that, in their featuring seconds, are mostly dissonant (from the viewpoint of Western music theory).

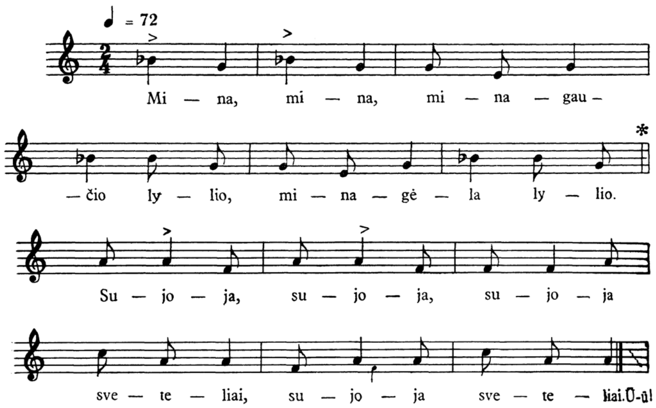


Figure 3. Transcription of one part of the *sutartinė* "Mina, mina, minagaučio lylio" (Slaviūnas, 1958, p. 657 [N428a]). The entrance of the canonically succeeding voice is asterisked.

Figure 3 shows a typical example of a *sutartinė*. Three singers canonically perform this *sutartinė* in such a way that the two parts (measures 1–6 and 7–12 in Figure 3) sound simultaneously, except in the beginning when only one voice (measures 1–6) sounds. The lyrics change. Thus mostly seconds occur continuously between the two voices.

Here I should note that because of the limitations of the five-lined staff, the transcription in Figure 3 is crudely misleading in terms of intervals (but still acceptable to comprehend the basics of performance). Actually, the intervals between the adjacent scale pitches do not split into two categories (whole tone and semitone), but rather fall into a single, relatively wide category centered at appr. 170-180 cents. We will return to this issue in the results section.

Sutartinės have only been registered in the northeastern part of Aukštaitija (see Figure 1) and only until the middle of the 20th century.

B. *Kanklės*

Kanklės is the same psaltery-type instrument as the Finnish *kantele*, the Estonian *kannel*, the Latvian *kokle*, and the Russian *gusli*. The trapezoid body of most *kanklės* is carved from a single piece of wood. The top plate attached to the body has resonance holes shaped mostly like stars ("suns") or other patterns that create small openings. Five or more gut or metal strings are tightened by pegs at the wide end and anchored to the metal bar at the narrow end. *Kanklės* strings are plucked with fingers of the right hand or with a pick. The left hand is used to stop the strings that should not sound and, sometimes, to pick additional notes. The *kanklės* is placed horizontally on the knees or somewhat higher and more vertically, or even on a table, for better resonance. Traditionally, only men played *kanklės*.

There are basically three types of *kanklės*. The smallest and most archaic instruments, mostly containing five strings were typical for northeastern Aukštaitija (Figure 4). *Sutartinės*-type music (Figure 5) was played by a single performer as a kind of meditation. Two other types of *kanklės*, found in northwestern Aukštaitija and Žemaitija, and in northwestern Žemaitija and Suvalkija, are larger and contain more strings, usually nine or twelve. They were presumably tuned diatonically, in a major-like scale, sometimes with additional basses. They were used not only for individual playing, but also for entertainment and for songs and dance accompaniments (Figure 6). Suvalkian *kanklės* are considered the most modern; they are sometimes richly decorated, with characteristic round and spiral endings (Figure 7).

Figure 4. *Kanklės*, northeastern Aukštaitija

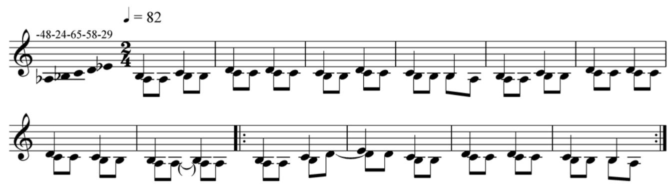



Figure 5. Transcription of the beginning of a *kanklės* piece "Sidijo, beržel, sidijo leliula" (Nakienė & Žarskienė, 2004, N5). Deviations in cents from the twelve-tone equal temperament are denoted

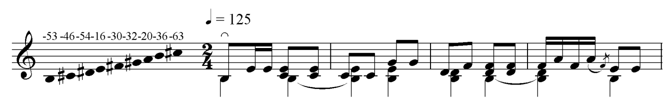



Figure 6. Transcription of the beginning of a *kanklės* dance piece "Kryžiokas" (Nakienė & Žarskienė, 2005, N19)

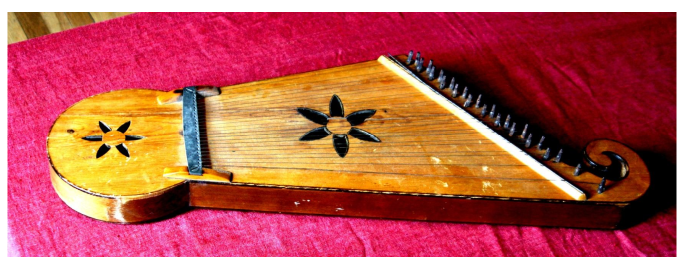


Figure 7. *Kanklės*, Suvalkija

II. ACOUSTIC MEASUREMENTS

A. Samples

Four samples were created for the analysis. The first and second samples contain 27 recordings of vocal *sutartinės* and 28 recordings of *kanklės* *sutartinės*-type music from northeastern Aukštaitija available at the Folklore Archives of the Institute of Lithuanian Literature and Folklore, Vilnius (also found at http://archyvas.llti.lt/irasai/). Some of the recordings are also published in a book and CD set by Nakienė & Žarskienė (2004).

Ten recordings of Suvalkian multipart homophonic singing comprising the third sample were used from a collection by Četkauskaitė, 2002. And finally, the fourth sample consists of 11 Suvalkian *kanklės* recordings also from the Folklore Archives of the Institute of Lithuanian Literature and Folklore; some of them are published in Nakienė & Žarskienė (2003).

B. Method

Software Praat was applied for the acoustic measurements. Frequencies of prominent fundamentals or other harmonics were measured from LTAS spectra of the recordings. The average frequencies were evaluated as Praat "gravity centers" of the corresponding filtered peaks of partials. Then the pitches were calculated employing the logarithmic frequency-pitch relation.

III. RESULTS: NORTHEASTERN SAMPLES

A. Singing

The measured scales of northeastern (Aukštaitian) polyphonic vocal performances (*sutartinės*) are presented in Figure 8. The vertical dotted lines separate four groups of singers; localizations (settlements) of the groups are specified below.

I should mention my earlier findings (e.g. Ambrazevičius, 2008) that the tonal structures of vocal *sutartinės* can be roughly considered as anchored on the tonal nuclei and dissipating, in terms of salience, towards the scale margins. The tonal nucleus can be approximated as "double tonic", i.e. containing two central pitches separated by the interval of a second. For instance, the most prominent and the most frequent dyad in the *sutartinė* "Mina, mina, minagaučio lylio" (Figure 3) is the one containing pitches A4 and Bb4; they form the tonal nucleus.

Therefore, to visualize somewhat the tonal structures, the scales in Figure 8 are normalized to the lower pitch of the tonal nucleus.

At first glance, the scales in Figure 8 show very different interval relations. However, one should notice that the intervals between the pitches of tonal nuclei (on the zero line and the one above) are very similar, even among different groups of singers. The average nuclear interval of the 27 examples is 181 cents; the standard deviation equals 11 cents. Other pitches are more freely intoned, which corresponds to their lower weights in the tonal structures. Additionally, in some of the examples these pitches occur too episodically to be reliable for evaluation.

B. *Kanklės*

Figures 9 and 10 show the measured tunings of five-string northeastern *kanklės*, both original and normalized to the lowest pitch. It would be problematic to normalize the tunings to the lower pitch of the tonal nucleus since, unlike the case of vocal *sutartinės*, the tonal nucleus is barely identifiable. Say that the general structure of the piece presumes that the nucleus should lie on the upper strings. The upper strings are damped when plucking the lower strings, but the lower strings are usually left to ring when upper ones are being plucked. Therefore, the resounding lowest strings often become the most important acoustically, thus masking the category of a tonal nucleus.

Of course, only the pitches used in the pieces can be measured, therefore some of the depicted tunings are incomplete. Since *kanklės* performer Plepas did not use the lowest pitch in pieces 13, 19, and 20, the relative pitches for these pieces are roughly defined with references to the adjacent pieces.

Performer Lapienė tuned the marginal strings in slightly stretched fifth (by 0–33 cents). The intervals between the adjacent pitches seem not to be very strictly defined, in terms of a whole tone/semitone, though some pattern of asymmetry is observed. Plepas tuned the marginal strings of his *kanklės* to a considerably larger interval, a stretched augmented fifth or even a minor sixth.

The pieces in Figures 9 and 10 are presented in chronological order, i.e. in the order they were recorded. Thus Lapienė's tuning rose slightly over time. Most probably this was not an intentional retuning but rather caused by increasing string tension due to decreasing temperature. However, starting from piece 8, at least the fourth string (counting from the lowest one) was either intentionally lowered or the musician did not notice the changed pitch of the loosened string.

Plepas' tunings show more distinct changes. It is important to note that the musician presents two different tuning procedures recorded as pieces 24 and 25. However, he considers the two tunings equally acceptable for playing the same music pieces. We are lucky that there are several recordings of the same pieces. Specifically, pieces 13 and 20; 14 and 21; 15, 17, and 26; 16, 18, 22, and 27; 19, 23, and 28 are actually multiple recordings of the same pieces. In addition, the first tuning (i.e. presented by pieces 13-24) shows significant variations from piece to piece.

The steadiest intervals are the ones between the two lowest pitches and the one above. However, even their standard deviations equal, respectively, 22 and 27 cents, i.e. they are considerable larger than in the case of the tonal nucleus in vocal *sutartinės*.

Briefly, a precise tuning seemingly was not of importance for the northeastern *kanklės* players. This is reflected in the freedom of Plepas' tunings and in the significant divergences of Lapienė's and Plepas' tunings used to perform music of the same type.

IV. RESULTS: SOUTHWESTERN SAMPLES

A. Singing

Similarly as with the NE sample, I have measured the scales of SW (Suvalkian) homophonic multipart vocal performances. This music has a more ordinary tonal structure than *sutartinės*. It is major-like. Therefore this time I normalized the scales to the tonic (Figure 11). Again, the vertical dotted lines separate the groups of singers.

Even though the sample is very small, it shows significant differences between the examples. Since this type of performance is based on sonorities in thirds (and not in seconds, as in NE case), the intonations of the corresponding dyads should be considered. So, the dyad I-III (tonic – third degree; here and further I use Roman numerals for the degrees) equals 422 cents, on the average, and the standard deviation is 20 cents. The corresponding numbers for the dyads II-IV, III-V, and IV-VI are 281 and 40 cents, 302 and 24 cents, 395 and 34 cents, respectively. The large standard deviation means that the scales are characteristic of relatively loose interval compositions.

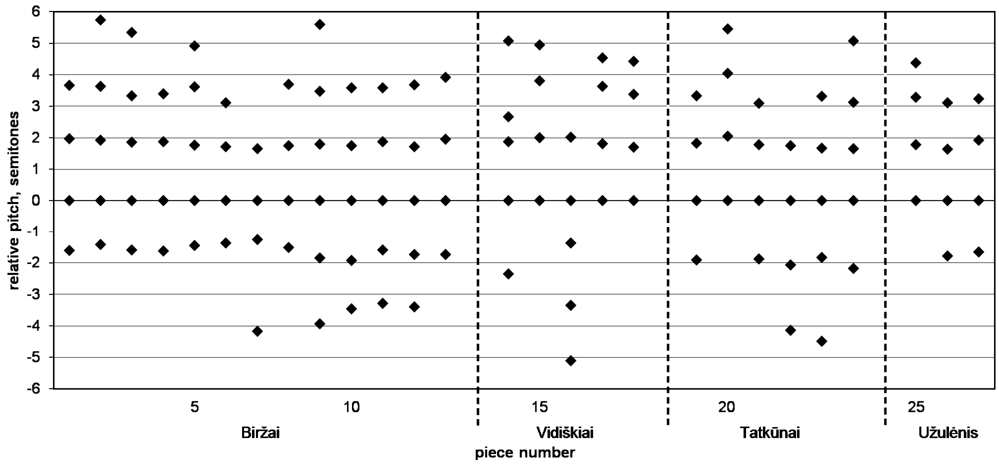


Figure 8. Scales of NE vocal performances (polyphony) normalized to the lower pitch of the tonal nucleus

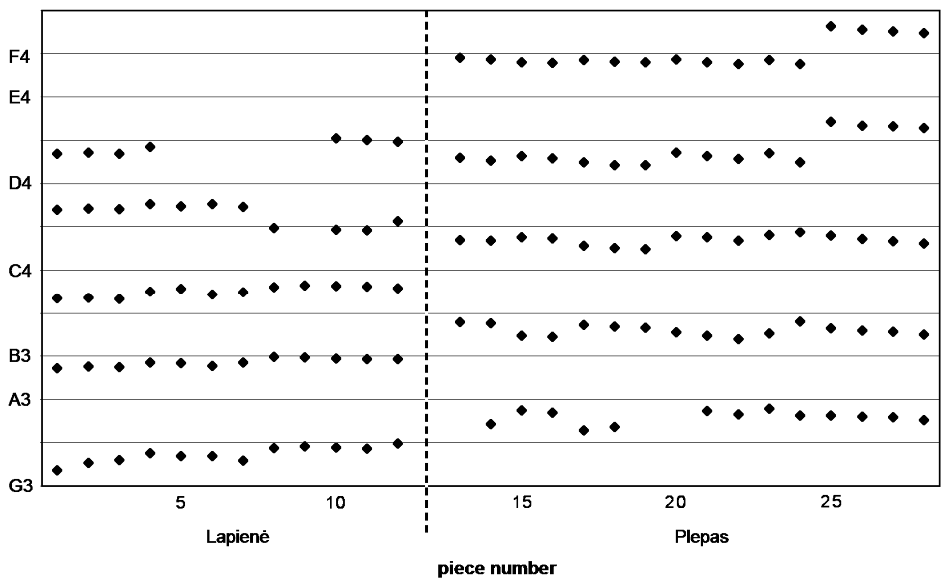


Figure 9. NE *kanklės* tunings

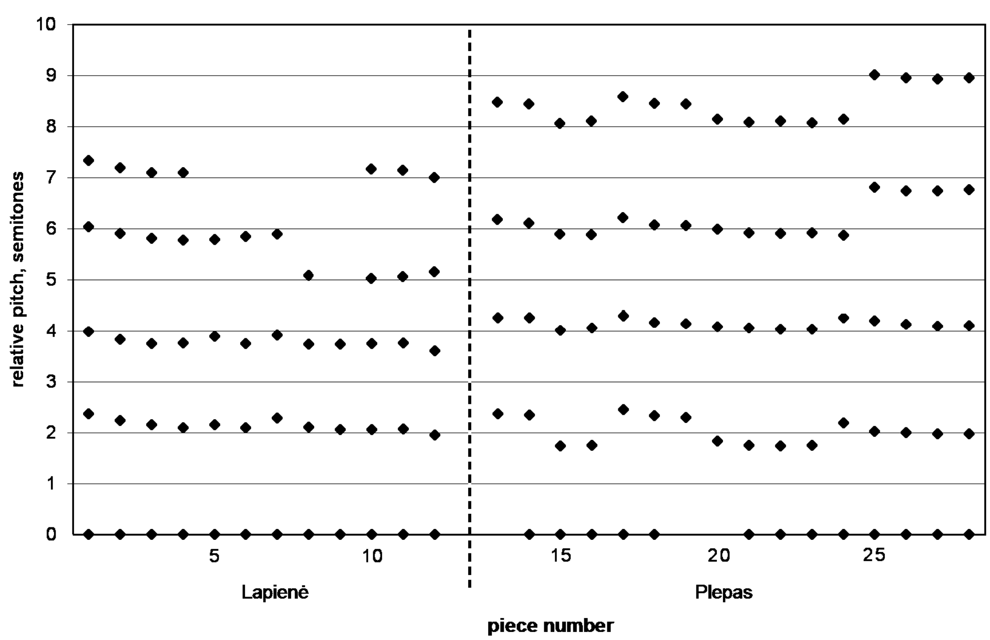


Figure 10. NE *kanklės* tunings normalized to the lowest pitch

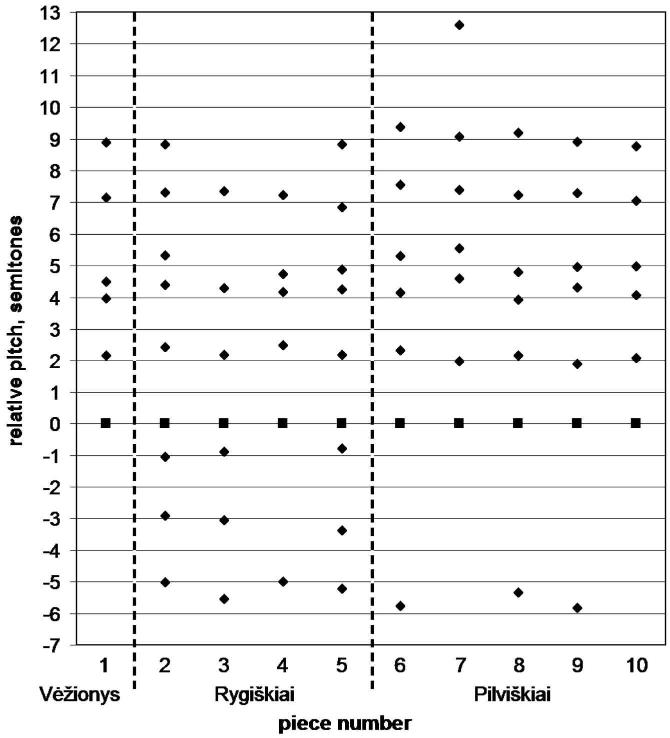


Figure 11. Scales of SW vocal performances (multipart homophony) normalized to the tonic (squares)

B. *Kanklės*

As was already mentioned, SW *kanklės* contain more strings, and the music is completely different, compared to NE *kanklės*. The tunings are major-like, as is most of the vocal music in this region. The measured *kanklės* tunings are presented in Figures 12 and 13.

We can state that the musicians were quite careful about tuning. This especially holds for Puskunigis: the tonic is somewhat raised or lowered in different pieces (Figure 12), while the rest of the pitches are raised or lowered, in fact, by precisely the same increment. This is clear from the normalized representation of the tunings (Figure 13).

As with the SW vocal performances, now we consider the intervals of thirds between the dyad pitches. At first glance, there is a noticeable difference between Popiera's and Puskunigis' tunings. However, a closer inspection reveals that, in the Popiera case, actually only the triad II-IV-VI is raised with no significant change of the intervals in the triad. The intervals between the triads I-III-V and II-IV-VI are of little importance as these triads or their separate pitches never sound simultaneously.

The pooled tunings of the two musicians show these averages and standard deviations: 408 and 7 cents (I-III), 300 and 7 cents (II-IV), 303 and 15 cents (III-V), and 398 and 11 cents (IV-VI). First, the numbers show relative precision in tuning (as opposed to the case of SW vocal performances), and second, a slight tendency to stretch the entire scale (similar to the case of SW vocal performances). This tendency stands out more clearly if the pitches an octave apart are compared: the interval between the fifth degrees above and below the tonic is 1228 cents, on the average.

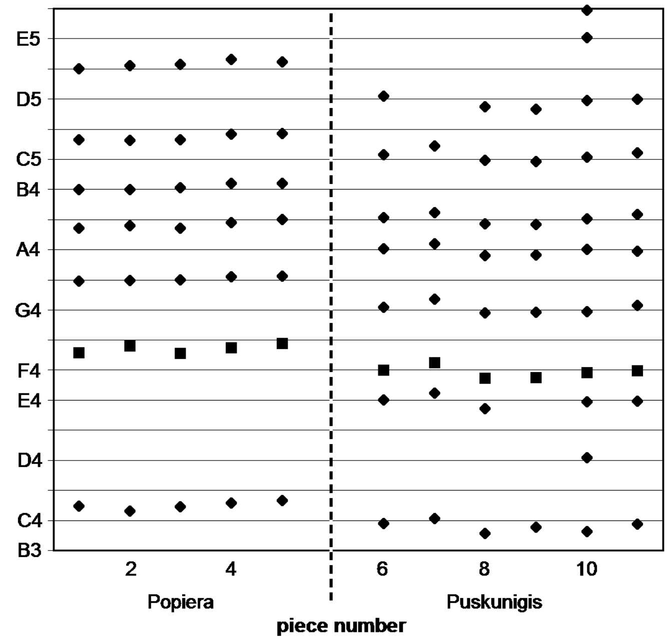


Figure 12. SW *kanklės* tunings. Squares denote tonics

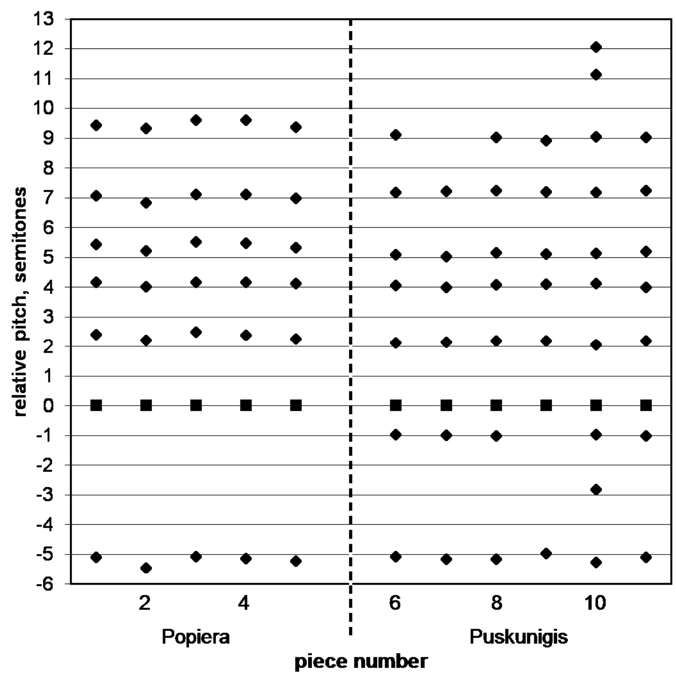


Figure 13. SW *kanklės* tunings normalized to the tonic

V. DISCUSSION

I have measured and analyzed musical scales in four samples of Lithuanian traditional vocal and instrumental music. The intent was to compare "looseness" of the scales in stylistically similar instrumental (*kanklės*) and vocal performances. I considered newer diatonic major-like melodies from southwestern part of Lithuania (Suvalkija) performed by singers and musicians. I found that *kanklės* tunings used for the melodies show relatively steady and similar patterns. The same or related melodies sung even by the same performers show more "loose" scales, which seems natural because of the differences in vocal / fixed instrumental intonation. The related sung melodies show more "loose" scales, which what seems natural because of the differences in vocal / fixed instrumental intonation.

However, a reverse tendency was found for the music from northeastern Lithuania (Aukštaitija). I considered only the more archaic music layer from this region: vocal *sutartinės* and *sutartinės*-type *kanklės* music. This style represents polyphony based on beating sonorities (*Schwebungsdiaphonie*, beat diaphony), i.e. harmonic intervals of seconds are characteristic. The basic dyads in vocal *sutartinės* tend to be quite steadily intoned. Supposedly, it is because the singers strive for maximum psychoacoustic roughness as for the ideal "clash" of voices. In earlier studies (e.g. Ambrazevičius, 2008), the maximum roughness for the frequency range of characteristic intense partials in the vocal spectra has been evaluated to occur at roughly 170-200 cents. Thus not surprisingly similar values for the basic dyads of *sutartinės* were registered in the present study as well. The other intervals in vocal *sutartinės* are more or less scattered, which is in agreement with less severe requirements of the "clash" for the marginal scale intervals.

Striving for maximum roughness seems to be less relevant in the *kanklės* case; the tunings are relatively "loose." I propose that this is because of the scale type and the playing technique. As mentioned, *sutartinės*-type music, both vocal and instrumental, belongs to a relatively archaic music layer that may be considered an instance of Alexeyev's "γ-intonation" (Alexeyev, 1986, p. 53). This type of scale is characterized by non-fixed or "wandering" pitches, wide pitch categories, and coordination of pitches, not subordination. Phylogenetically the γ-intonation is a precedent to the modern type of scale with relatively fixed pitches, narrow pitch categories, and the subordination of pitches (Alexeyev's "t-intonation").

In the case of vocal *sutartinės*, the "wandering" pitches are noticeably stabilized by the requirement to produce psychoacoustic roughness, whereas the sustained rough dyads or clusters producing roughness cannot be realized by relatively fast decaying and usually shortened (stopped) *kanklės* sounds.

I should stress that though the samples are taken from different ethnographical regions, the primary focus is not on regions, but on styles. That is, I compare cases of homophony and a peculiar type of polyphony. Some pilot analysis shows that the newer homophonic multipart singing in northeastern Aukštaitija is similar to the examined homophonic multipart singing in Suvalkija in terms of scale "looseness." Hence, the discussed traits of scales should be first attributed to style, not to musical dialect.

In addition, solo singing and homophonic multipart singing from different musical dialects are often characteristic of gradual tonality change, mostly a rise of the entire scale from the beginning to the end of performance. The gradual change of scale position is usually supplemented with modification of the intervals between scale degrees (Ambrazevičius, 2014). However, the change of intervals does not affect the percept of scale: emically, from the performers' viewpoint, the scale is considered to remain constant. These gradual changes point at certain aspects of the discussed "looseness." Importantly, the gradual scale changes are not observed in the case of vocal *sutartinės*; the scales are acoustically stable in the course of performance.

In the present study, I employed music examples from two "instruments" in a broad sense, — *kanklės* and voice — to demonstrate the examined phenomena. It is possible to extend the study to include other instruments. Some aspects of scale "looseness" in the cases of other Lithuanian traditional string and wind instruments are discussed in Ambrazevičius, Budrys, & Višnevska (2015).

VI. CONCLUSION

The "looseness" of scales results from a number of factors. Specifically, for the examined examples of traditional music, it was demonstrated how the phylogenetically determined type of scale, requirements of psychoacoustic roughness, and constraints of performance technique can combine and produce various instances of the "looseness."

REFERENCES

Alexeyev, E. (1976). *Проблемы формирования лада [Problems of modality development]*. Москва: Музыка.

Alexeyev, E. (1986). *Раннефольклорное интонирование. Звуковысотный аспект [Early folklore intonation: The pitch aspect]*. Москва: Советский композитор.

Ambrazevičius, R. (2008). Psychoacoustical and cognitive basis of Sutartinės. In K. Miyazaki, Y. Hiraga, M. Adachi, Y. Nakajima, Ų M. Tsuzaki (Eds.), *ICMPC10. Proceedings of the 10th International Conference on Music Perception and Cognition. 25–29 August 2008. Sapporo, Japan* (pp. 700-704). Adelaide: Causal Productions.

Ambrazevičius, R. (2014). Performance of musical scale in traditional vocal homophony: Lithuanian examples. *Музикологија / Musicology, 17*, 45-68.

Ambrazevičius, R., Budrys, R., & Višnevska, I. (2015). *Scales in Lithuanian traditional music: Acoustics, cognition, and contexts*. Kaunas: Kaunas University of Technology, ARX Reklama.

Ambrazevičius, R., & Wiśniewska, I. (2008). Chromaticisms or performance rules? Evidence from traditional singing. *Journal of Interdisciplinary Music Studies, 2*(1&2), 19-31.

Četkauskaitė, G. (Ed.) (2002). *Lietuvių liaudies muzika III d. Suvalkiečių dainos. Pietvakarių Lietuva [Lithuanian folk music. Vol. 3. Suvalkian songs. Southwestern Lithuania]*. Vilnius: Lietuvos muzikos akademijos Muzikologijos instituto Etnomuzikologijos skyrius.

Fastl, H., & Zwicker, E. (2007). *Psychoacoustics. Facts and models*. Berlin, Heidelberg, New York: Springer.

Friberg, A., Bresin, R., & Sundberg, J. (2006). Overview of the KTH Rule System for musical performance. *Advances in Cognitive Psychology, 2*(2&3), 145-161.

Fyk, J. (1994). Static and dynamic model of musical intonation. In A. Friberg, J. Iwarsson, E. Jansson, & J. Sundberg (Eds.), *Proceedings of the Stockholm Music Acoustics Conference. July 28-August 1, 1993* (pp. 89-95). Stockholm: Stockholm: Royal Swedish Academy of Music.

Hess, W. (1983). *Pitch determination of speech signals. Algorithms and devices*. Berlin: Springer Verlag.

Kopiez, R. (2003). Intonation of harmonic intervals: Adaptability of expert musicians to Equal Temperament and Just Intonation. *Music Perception, 20*(4), 383-410.

Nakienė, D., & Žarskienė, R. (Eds.) (2003). *Suvalkijos dainos ir muzika. 1935–1939 metų fonografo įrašai / Songs and music from Suvalkija. Phonograph records of 1935–1939*. Vilnius: Lietuvių literatūros ir tautosakos institutas.

Nakienė, D., & Žarskienė, R. (Eds.) (2004). *Aukštaitijos dainos, sutartinės ir instrumentinė muzika. 1935–1941 metų fonografo įrašai / Songs, Sutartinės and music from Aukštaitija. Phonograph records of 1935–1941*. Vilnius: Lietuvių literatūros ir tautosakos institutas.

Nakienė, D., & Žarskienė, R. (Eds.) (2005). *Žemaitijos dainos ir muzika. 1935–1941 metų fonografo įrašai / Songs and music from*

Žemaitija. Phonograph records of 1935–1941. Vilnius: Lietuvių literatūros ir tautosakos institutas.

Parncutt, R., & Cohen, A. J. (1995). Identification of microtonal melodies: Effects of scale-step size, serial order, and training. *Perception and Psychophysics, 57*(6), 835-846.

Slaviūnas, Z. (1958). *Sutartinės. Daugiabalsės lietuvių liaudies dainos [Sutartinės. Lithuanian multipart folk songs]*. V. 1. Vilnius: Valstybinė grožinės literatūros leidykla.

Effect of Instrumental Studies on Pitch Labeling

Christopher Bartlette[1]
[1]Department of Music, Binghamton University, Binghamton, NY, USA

Persons who do not possess absolute pitch (AP) often have good pitch memory—for example, the ability to recognize the pitch level of a familiar song or a tone they have heard many times (such as an orchestral tuning pitch). The goal of this study is to investigate whether playing a musical instrument provides a context not just for pitch memory, but also for labeling pitches. Because persons who play instruments receive physical, visual, and aural cues that connect with specific pitches, it is expected that the reference to a musical instrument will increase the accuracy in labeling pitches. In a first experiment, 12 participants—persons who played non-transposing instruments and did not possess AP—were asked (without any context) to move a slide on a tone generator until the tone matched the pitch C (in any octave). After a brief discussion, during which participants' information was gathered, the participants were asked to envision themselves playing their instrument and stopping on the pitch C. Then, the participants moved the tone generator until it matched the imagined instrumental pitch C. The data were recorded as distances from the nearest pitch C, from 0 to 600 cents (0 to 6 semitones). A paired-sample t test found that the imagined instrumental pitches were nearer to C ($M = 120¢$) than the pitches without context ($M = 225¢$), $p = .02$. In a second experiment, 50 participants—persons who play non-transposing or transposing instruments—complete the same tasks. It is expected that those who play non-transposing instruments will replicate the results of the first experiment, and those who play transposing instruments will match the pitch C written for their instrument, rather than the pitch C written for non-transposing instruments. Data collection is underway for the second experiment. The results of these experiments will be discussed within the context of AP, as well as pedagogical issues with teaching aural skills to instrumentalists.

ISBN 1-876346-65-5 © ICMPC14

Psychophysical and Electrophysiological Measures of Auditory Imagination, Expectation, and Creativity in Jazz Musicians

Emily Przysinda[1], Kellyn Maves[1], Cameron Arkin[1] Tima Zeng[1], Pheeroan Aklaff[2], Psyche Loui[1]
[1]Department of Psychology and Program in Neuroscience and Behavior, Wesleyan University, Middletown, CT, USA
[2]Department of Music, Wesleyan University USA

Are Jazz musicians more creative? Here we use domain-specific and domain-general measures of creativity and imagination to investigate musical creativity in Jazz musicians, Non-jazz (mostly classical) musicians, and Non-musicians. Our test battery includes measures of auditory imagery, expectation, and divergent thinking skills. The scale imagery task (modified from Navarro-Cebrian & Janata, 2010) had participants listen to scales and judge whether the last note was modified in pitch. Trials included a perceptual condition, where all notes play, and an imagery condition, where some of the notes are left out. Results showed that while all subjects performed above chance, Jazz musicians were more accurate than Non-musicians in musical scale perception as well as imagery. Psychometric functions for Jazz and Non-jazz musicians showed steeper slopes compared to Non-musicians. In the harmonic expectation task, participants rated their preference of chord progressions of different expectancies while EEG was recorded. Behavioral data showed that Jazz musicians preferred slightly unexpected chords over highly expected chords, and were also more tolerant of very unexpected chords. ERP data showed a more prominent P300 and early right anterior negativity in the highly unexpected chords for the Jazz musicians compared to the Non-jazz musicians and Non-musicians. This suggests that Jazz musicians are more sensitive to unexpected musical events. The Divergent Thinking task (Torrance 1968), which required participants to respond to prompts in a creative manner, was used as a domain-general measure of creativity. There was a main effect of group for four out of the six questions in fluency and originality measures, with the Jazz musicians scoring highest. Together, results suggest that musical experience, especially in Jazz musicians, confers an advantage in auditory imagery, musical expectancy, and domain-general creativity, which may be useful in creative improvisation.

Manipulating Melodic Expectation

Jason C. Rosenberg[1], Allison R. Fogel[2], and Aniruddh D. Patel[2]
[1]Center for New Music, San Francisco, CA, USA
[2]Psychology Department, Tufts University, Medford, MA, USA

The melodic cloze probability method (Fogel et al., 2015, *Front. in Psychol.*) uses a production task to quantitatively measure melodic expectation: listeners hear the opening of a novel, monophonic melody (i.e., a melody without any accompaniment) and sing the note they think comes next. By manipulating the structure of the melodic openings, one can study how different structural factors interact in shaping melodic expectation. The current study examines how subtle changes to melodic structure influence expectation for the tonic note. 48 melodic openings (each 5-9 notes in length) were composed with the aim of creating expectation for the tonic using an implied authentic cadence. Participants were asked to sing a single note continuation of each melody. In 26 of these melodies, the large majority of participants did in fact sing the tonic, but in 22 of the melodies, less than 70% of participants sang the tonic (mean cloze probability of tonic = 55%). These 22 melodies were subtly modified to increase expectancy for the tonic, and were tested on a new set of participants. In the revised melodies, on average the tonic was sung more frequently (mean melodic cloze probability of the tonic increased from 55% to 78%). The structural factors that most shaped the increase in melodic expectation include not only local pre-target changes (such as reducing intervallic distance, changing the melodic patterning, or adding a neighbor gesture around target), but also upstream changes (such as rhythm simplification or elongation, or changes to implied harmony). However, not all revised melodies resulted in our goal of attaining a significant increase in melodic expectation for the tonic. Analysis of the successful and failed attempts at manipulating melodic constraint within cadential sequences reveal an interplay of local and global factors in shaping melodic expectation.

Can I Have the Keys?: Key Validation Using a MIDI Database

Joshua Albrecht,*[1] Daniel Shanahan#[2]

*Music Department, The University of Mary Hardin-Baylor, Belton, TX, USA

#School of Music, Louisiana State University, Baton Rouge, LA, USA

[1]jalbrecht@umhb.edu, [2]dshanahan@lsu.com

ABSTRACT

The accuracy rates of a recently-developed key finding algorithm were estimated to be quite high, higher than most other currently available algorithms. This study provides a) an external validation of the accuracy rates of that algorithm on an independent dataset, and b) a validation of that dataset using the algorithm. Specifically, a large dataset of over 12,000 MIDI works found on ClassicalArchives.com were sampled for this purpose. MIDI data tends to contain a high percentage of errors. To determine whether the database provides a sufficient signal to noise ratio for pitch-reliant tasks like key-finding, an initial validation study was performed on the dataset in which accuracy rates for the key-finding algorithm on a small subset of 32 works were compared with a ground truth of accuracy rates. In the initial study, it was found that 28 of 32 (87.5%) of MIDI works returned the same result as the ground truth. Given this relatively high accuracy rate, it was determined that the database can be an appropriate vehicle for testing a key finding algorithm. On a dataset of 192 separate works, it was found that the algorithm correctly assigned keys to 164 works (85.4%). Together, the two studies are consistent with the ideas that, 1) though there will be some errors as a result of noisy data, a MIDI database can serve some uses for some studies, and 2) the key finding meta-algorithm provides an accurate way to assign keys to this database.

I. INTRODUCTION

MIDI data can be notoriously messy. In an unpublished validation study, Shanahan and Albrecht (MS) found that ClassicalArchives.com, consisting of a large MIDI database of over 12,000 user-generated works, was exceptionally prone to errors. They measured fifteen different types of error, including pitch, rhythms, meter, and key signature error. Depending on the type of error, error rates ranged from a low of 5.18% (± ~3%) for misaligned note onsets to a high of 53.88% (± ~7%) for incorrect number of measures due to blank measures opening files.

Despite the untrustworthy nature of the data, the dataset is nevertheless an attractive potential source of future research. There are already studies beginning to be published that draw on this particular dataset (e.g. White 2013, 2014). One reason this large database is attractive is because most of the current digitally encoded datasets available have been used extensively in many studies for years. The overuse of particular datasets in empirical studies could lead to methodological errors such as double-use data or multiple tests, and this new repository offers a way to avoid using the same data too many times. A second reason for using this dataset is because it is more representative of composers and ensemble sizes than most currently available datasets, which tend to focus on Germanic composers and small ensembles.

Given the advantages offered by adding such a database to the available resources, it is reasonable to further investigate its viability for different types of research questions. For example, even if the error rates are high for note durations, one might ask if reliable estimates of scale degree distributions can still be made.

For the purposes of illustration, consider Figure 1. This figure shows an encoding from the database for the first seven measures of Chopin's Nocture in G minor, Op. 15, No. 3. The top of Figure 1 is from the 1915 Rafael Josephy Schirmer edition. Underneath this excerpt is a visualization of the MIDI data, obtained by converting the MIDI data to MusicXML and opening it in Musescore 2.0.2. Inspection of the two versions reveals sizable differences in metric placement, rhythmic durations, and number of note attacks, along with missing dynamics and articulations.

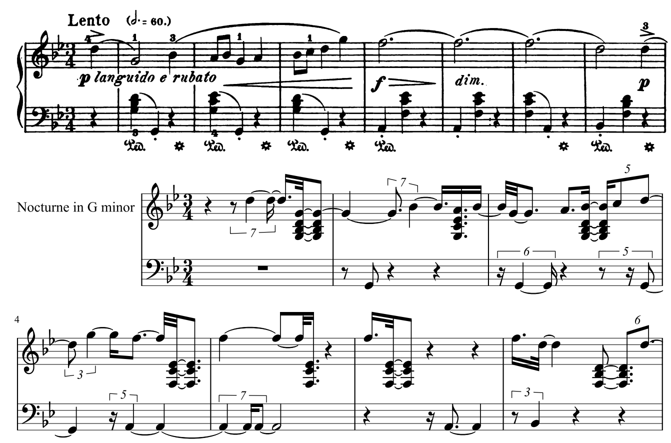



Figure 1. An illustration of some of the problems of MIDI data. The top image is from a 1915 published score of Chopin Op. 15, No. 3 and the bottom is a visual representation of corresponding MIDI data. As can be seen, there are many discrepancies.

Another obstacle for the practical use of MIDI data is a tendency for no key interpretation to be provided. That is, pitch-classes are present with a key signature in more than half of the cases, but most files do not explicitly identify the key. Compounding this issue is a tendency for provided key signatures to be incorrect (~40% of the time). While certain questions can be answered without key identifications, any attempts to analyze functional progression, scale degree tendencies, or other musical elements related to key are greatly hindered without this information.

Even considering the gross inaccuracies that tend to plague this data, most of the pitch information is relatively intact. Shanahan and Albrecht (MS) estimate that only about 8.5% (± ~2%) of pitch information in this database is incorrectly encoded. It may be the case, then, that this dataset may be equipped to answer certain types of questions involving pitch.

One such question is the accuracy of key finding algorithms, which is a particularly germane issue for three primary reasons. First, key finding algorithms only consider distributions of pitch-classes. Although there are some pitch-class encoding issues in the dataset, accuracy rates are relatively high. Moreover, music theorists have historically focused on pitch information more than on other types of musical structures for analysis. Therefore, resolving whether or not the database can reliably answer pitch-related questions promises to offer the most diverse analytical benefits.

Second, as mentioned above, one way to greatly increase the database's usefulness for examining functional or scale-degree information would be to assign keys to each file. By testing to what extent key finding algorithms perform accurately on this dataset, we can move towards using the database more constructively.

Finally, key finding algorithms have undergone some significant changes in recent years. Specifically, refinements and improvements have been made to key finding strategies, resulting in higher accuracy rates. One such algorithm uses Euclidean distance to estimate keys, rather than the more common use of correlation (Albrecht & Shanahan, 2013). In their study, Albrecht and Shanahan obtained an accuracy rate of 93.1% on a database of 982 works encoded in humdrum kern notation.

By combining their algorithm with the Aarden-Essen algorithm (Aarden 2003), they were able to create a meta-algorithm that performed at 95.1% accuracy. However, because the algorithm was determined in a *post hoc* manner, further research on an independent dataset is needed to fully assess the actual accuracy rate. The ClassicalArchives.com dataset provides such an opportunity to test this algorithm's veracity.

II. AIMS

In light of the above considerations, this study has two complementary aims: 1) we will examine how appropriate the classical archives MIDI dataset is for examining pitch-related inquiries like key profiling and key finding, and 2) we will test the Albrecht & Shanahan (2013) meta-algorithm on this independent dataset to determine how accurate the algorithm is. If the results are consistent with both the idea that the MIDI dataset can be used for certain pitch-related tasks and that the meta-algorithm is as accurate as Albrecht & Shanahan (2013) suggest, then using this new dataset with algorithmically assigned keys can open up new possibilities for computational research.

III. METHODOLOGY

A. Sampling

The ClassicalArchives.com dataset is relatively large by the standards of classical music corpora. At over 12,000 files, the database promises to significantly augment current digitalized offerings. However, validation studies tend to be very labor intensive, requiring visual inspection of individual files. As a result, some means of sampling a small subset of these files is necessary.

Recall that there are two primary goals for this study. The first goal is to assess the viability of using this relatively messy data for key finding. The second goal is to test the key-finding algorithm. Different subsets are required for each individual goal.

1. Database validation sampling

The first step we used in the process of sampling for these two goals was simply to excerpt roughly 10% of the dataset. We first sampled 1,200 files from the dataset. After the conversion process, there were issues with several of the files (see III: B below). These problematic files were discarded, resulting in 1,125 files.

For the purpose of assessing how viable these data are for a key finding task, some ground truth was required. Therefore, the methodological decision was made to pair files from the 1,125 file subset with the corresponding musical works used in Albrecht and Shanahan (2013). In that study, 982 works had previously had keys assigned by independent analysts. These works had been encoded into the Humdrum kern representation, and the assigned keys were used as a ground truth for the key finding algorithm. By matching files from the current study's subset to the works used in that study, a ground truth would enable the testing of this dataset.

Importantly, note that the goal in the first phase of sampling is not to determine how accurate the key finding algorithm is. Instead, we wish to assess *how similarly the algorithm performs* on these messy MIDI files as it did on that study's cleanly-encoded files. Because we already know how well the algorithm performs on clean data, we can determine whether the algorithm performs in a similar or different way on MIDI data. In short, we will not be comparing the algorithm's results on the MIDI to the ground truth of analyzed keys, but rather to the ground truth of the algorithm's results on kern data. Therefore, it is important that we only use the exact same movements or works used in the first study.

After matching works used in Albrecht and Shanahan (2013) to the subset of 1,125 files selected for this study, only 32 files were found in common between the two datasets. These 32 files were used as a comparison to determine the effect of encoding on algorithm performance. Members of this subset included Bach's *Well-Tempered Clavier* and Brandenburg concerto movements, Beethoven piano sonata movements, Beethoven, Mozart, and Haydn string quartet movements, Scarlatti piano sonatas, Chopin Op. 28 preludes, Hummel Op. 67 *Préludes*, and movements from Vivaldi's *The Four Seasons*.

2. Key finding algorithm validation sampling

Depending on the results from the first element of this study, it may be appropriate to validate the key finding algorithm on the MIDI dataset. Of course, this depends on the initial results. If the way that the algorithm performs on the MIDI set of 32 files is significantly different from the way it performs on the Humdrum files, then that reveals that this dataset is likely inaccurate enough to be ill equipped for key finding tasks.

Each file used in the key finding validation needs to have a key manually encoded by an analyst. This is a time consuming process, requiring visual inspection not of the MIDI file, but of an original publication of the composition or movement (in case of encoded errors in the MIDI file). Of course, the bigger the sample, the more confidence can be placed in the results from the study. As a middle-ground between these two values, 204 of the 1,125 files originally subsetted that were not included in the first sample were chosen for validation of the algorithm.

B.File conversion

The key-finding algorithm does not work directly on MIDI data, so in order for the analysis to be performed the data first needed to be converted into Humdrum's kern representation. However, tools that directly convert MIDI data to kern representations tend to perform poorly.

An alternative strategy is to use a middle-ground. Most computer-based music notation programs directly engage with MIDI data. For this study, we used MuseScore 2.0.2 to read the MIDI data.

When reading data from MIDI into MuseScore, shortest note length must be specified. To capture the most accurate representation of the data, we used the shortest note value of a 64th note. Once the data was read by MuseScore, it could easily be exported into MusicXML. From MusicXML, the conversion into kern format is relatively straightforward using the Humdrum extra tool xlm2hum, designed by Craig Sapp (2010).

Of the 1,200 works originally sampled for the subset, there were errors in at least one step of the file conversion process for 75 files. These 75 files were discarded from further consideration.

C.Key analysis

For the second stage of this study, files were required to be analyzed with key identifications as a ground truth to compare performance of the key finding algorithm against. Because some of the MIDI files contained a large number of errors, the decision was made to assign keys based on published editions of each work rather than on the MIDI files.

One difficulty in identifying the compositions linked to the files in the ClassicalArchives.com database is that the files do not always contain metadata. However, the website contains a directory of its musical works in html documents. The first step in identifying keys, then, was to identify the works. The html files were searched for the file names in the subset of 1,125 works.

Once the composer and work were identified, scores were downloaded from the International Music Score Library Project (imslp.org). Editions were randomly sampled, with the exception that recent creative commons publications were excluded. The reason why this exclusion criterion was established is because many of these files were generated from online encodings of works. Because our goal was to validate these online encodings against publications, scores generated from the online encodings were problematic.

The 204 works selected for validating the key finding algorithm were randomly assigned to one of the two authors of this study. The pdf documents were then analyzed for keys by one of the two authors. If the work was post-tonal, it was discarded as not appropriate for use with a key finding algorithm. Pre-tonal works were included if the tonal center and mode ('majorish' or 'minorish') was clear, otherwise they were discarded. Some of the MIDI files were also clearly transposed from their originally published versions; for the purpose of clarity, these were also discarded. Finally, works that began and ended in different keys were discarded from further consideration. After discarding these files, 192 works remained.

IV. RESULTS

A.Database validation

Recall that the initial goal of this study was to assess whether or not the MIDI data was viable for pitch-based tasks like key finding. In order to determine whether this was the case, we compared the key finding meta-algorithm's results on the 32 kern files used for ground truth to the 32 MIDI works sampled for this purpose, described in III: A, 1.

For example, assume that the algorithm estimates the keys of four works encoded in Humdrum format. For the sake of argument, we will call them compositions A, B, C, and D. If the algorithm correctly estimates the keys of A and B and incorrectly estimates the keys of C and D, the algorithm will have an accuracy rate of 50%. Once we locate MIDI versions of compositions A, B, C, and D, we can compare the algorithm's performance on the MIDI and Humdrum formats. If the same algorithm correctly estimates the keys on the MIDI versions of C and D and incorrectly estimates the keys of A and B, the accuracy will be 50% compared to the correct keys.

However, the performance of the algorithm on the MIDI data compared to the Humdrum data will be 0%. In short, because they are the same compositions, the algorithm should perform in the exact same way on either version of the data *if the representations of the data are equally useful for key estimation in either representation.* Because the algorithm performs differently on each work in this hypothetical example, we can safely conclude that the representation of the data between the two encodings has a large impact on key estimation.

Of the 32 kern files used from Albrecht and Shanahan (2013), 31 works were correctly assigned keys by their meta-algorithm (96.9%). Of the MIDI files of the matched works, only 29 works were correctly assigned keys (90.6%). Moreover, the one kern file incorrectly assigned a key, the first movement from Beethoven's string quartet No. 14, was *correctly* assigned to the corresponding MIDI file. Of the 32 works, then, only 28 were assigned matching keys, or 87.5%.

These results are consistent with the hypothesis that the errors involved in the MIDI data have an effect on certain pitch-based tasks, like key estimation. Nevertheless, 87.5% is not too different from the accuracy rates of many key finding algorithms. Although in this case the same algorithm is being applied to different formats of (allegedly) the same work, and so should be held to a higher standard, the results are nevertheless not out of the realm of normal performance. Therefore, we felt that it was appropriate to continue the validation of the algorithm. Nevertheless, it is important to keep in mind that at least some of the error in the following section may be attributable to error in the dataset rather than error in the key finding meta-algorithm.

B.Meta-algorithm validation

After it was determined that the MIDI database provided a reasonably satisfactory source for pitch class information, at least for the purposes of estimating keys using pitch-class distributions, the accuracy of the Albrecht-Shanahan meta-algorithm (2013) was tested against a sample of 204 files. As mentioned above, we established ground truth for keys by a visual inspection of the corresponding scores downloaded from the International Music Score Library Project (imslp.org). Works in which the key was not clear or where the MIDI file was clearly transposed from the original source were discarded.

Of the 192 files remaining, the key finding algorithm correctly identified the keys of 164 files, or 85.4%. Four of these files had the correct pitch-class assigned as a tonal center, but it was enharmonically respelled, such as a work in D$^{\flat}$ minor being identified as C# minor. The reason for these misattributions is likely due to the fact that the key finding algorithm picks an enharmonic key based on which accidentals are more prevalent in the piece. Because MIDI data only includes pitch-class information, MIDI files with no key signature or incorrect key signature (again, about 40% of the time (Shanahan & Albrecht, MS)) are interpreted by MuseScore as the pitch-class that appears more often in the repertoire. For example, B$^{\flat}$ is chosen over A# because it is a more common notated pitch. Because this lack of information is not an error with the algorithm, enharmonically respelled tonal centers were counted as matches.

Of the errors in key estimation, the most common error was assignment of the parallel key, occurring as 46.4% of errors. This result was also prevalent in the initial study (Albrecht & Shanahan, 2013). In this database, this error comprises a bigger percentage of the total. Many of these errors involved Baroque minor mode works. This is not surprising given the tendency of these works to employ a Picardy ending. This particular error, then, appears to be the result of the algorithm's focus on the first and last 8 measures of each work (for further details on how the algorithm works, see Albrecht & Shanahan, 2013). The second and third most common errors involved the relative key and fifth-related, each accounting for 25% of errors. The different types of errors and their frequencies are shown in Table 1.

Table 1. Types of errors in key-assignment. The most common type of error is assignment of the relative key.

Error type	# of occurrences (%)
Parallel key	13 (46.4%)
Relative key	7 (25%)
Fifth-related	7 (25%)
Non-relative mediant-related	1 (3.5%)
Step-related	1 (3.5%)

V. DISCUSSION

The results from this test of the algorithm reveal a lower accuracy rating than the initial study. This is not surprising, though, for several reasons. In the first case, the meta-algorithm was developed *post hoc* over several iterations to produce that formulation that would perform as well possible on the given data. It is likely, therefore, that there was some overfitting involved in the initial study. The current study is a direct consequence of the effort to test the extent of the overfitting involved in the first study.

In the second case, the lower accuracy ratings are unsurprising given the nature of the data. Because the MIDI data are fairly error-prone, a number of the mistakes in key assignment may be attributable to noisy data. Given the relatively high rate of errors in pitch-class (~8.5%) and rhythmic duration (~32%), both of which types of data are important in the estimation of key with this algorithm, it is not surprising that accuracy rates would be lower than in clean datasets.

Nevertheless, if a more generous calculation of accuracy is used, in which only the correct assignment of tonal center is checked, then this algorithm performs at a fairly impressive 92.2% accuracy rate, not significantly different from the accuracy rate found in Albrecht & Shanahan (2013).

Given the fluid nature of modality, especially in early common-practice era music and late common-practice era music, the identification of tonal center alone may be all that is possible in these styles. Because the ClassicalArchives.com dataset is broader than the sample used in the original study, it is possible that these early and late works are skewing the results of the study. In other words, it is possible that the accuracy rates are different because the samples represent slightly different populations.

Nevertheless, correct assessment of tonal center is still enough information to assign scale-degrees and harmonic function of chords. Although completely accurate key assignments of both tonal center and mode would be more useful, tonal center alone is enough to begin functional analysis.

These results are consistent with a) the hypothesis that the key-finding meta-algorithm is a reliable indicator of key, and b) that the MIDI database can somewhat reliably be used for the right pitch-related research questions. With the error rate so high in the comparison between the 32 files used in the original study and this study, it is difficult to determine the contribution of encoding error on key estimation error.

Future work will attempt to further disentangle these effects by further validating both the database and the meta-algorithm. Moreover, future research questions examined with this large and exciting new dataset will likely provide further insights into the nature of the data and its reliability.

REFERENCES

Aarden, B. (2003). *Dynamic melodic expectancy* (Unpublished doctoral dissertation). Ohio State University, Columbus, OH.

Albrecht, J., & Shanahan, D. (2013). The use of large corpora to train a new type of key-finding algorithm: An improved treatment of the minor mode. *Music Perception: An Interdisciplinary Journal, 31*(1), 59-67.

Sapp, C. (2010.). http://extras.humdrum.org/man/xml2hum/. Accessed 5/6/2016.

White, C. (2014). Changing styles, changing corpora, changing tonal melodies. *Music Perception: An Interdisciplinary Journal, 31*(3), 244-253.

White, C. (2013). An alphabet reduction algorithm for chordal n-grams. In J. Yust & J. Wild, *Mathematics and Computation in Music.* Heidelberg: Springer.

Effects of Familiarity, Key Membership, and Interval Size on Perceiving Wrong Notes in Melodies

Rachna Raman,*[1] Kevin Herndon,*[2] W. Jay Dowling*[3]

*School of Behavioral & Brain Sciences, The University of Texas at Dallas, USA
[1]rachna.raman@utdallas.edu, [2]kxh140630@utdallas.edu, [3]jdowling@utdallas.edu

ABSTRACT

Previous investigations showed that Western participants' perception of wrong notes in familiar Western melodies was influenced primarily by key membership (diatonic or nondiatonic) and to a lesser extent by interval size (1 or 2 semitones away from the original note). These results were supported by a cross-cultural study between South Indian classical (Carnātic) and Western musicians and nonmusicians with highly familiar Western melodies. However, Indian participants were slower and less accurate with Carnātic than with Western melodies, presumably due to the complexity of the Carnātic music system. In this study, we examined the effect of song familiarity in Westerners' perception of wrong notes. We chose 32 Western melodies previously rated as familiar, and 32 highly unfamiliar melodies, similar in style to the familiar melodies. Participants heard each melody twice, each time with one wrong note that was determined from one of eight possible types of wrong note based on key membership, interval size, and direction (up vs. down). Participants identified the wrong note by pressing a key. The results indicated an effect of music experience with unfamiliar melodies only, with musicians detecting wrong notes faster than nonmusicians. All groups were faster with familiar melodies than with unfamiliar melodies. Key membership influenced perception of wrong notes in both familiar and unfamiliar melodies: Participants were slower at recognizing wrong notes that were diatonic and faster when they were nondiatonic. Interval size influenced perception of wrong notes only in unfamiliar melodies: Participants were slower at recognizing wrong notes that were 2 semitones away than when they were 1 semitone away, which was the opposite of our previous studies with familiar melodies. We take these results as converging evidence that people remember the pitches in melodies in terms of steps of an overlearned modal scale, and not as successive intervals.

INTRODUCTION

The melodic-rhythmic contour of a melody—the pattern of ups and downs of pitch along with the relative durations of the notes—is an important feature, both in terms of musical structure and of memory. But there is more to a melody than just contour; the pitch intervals among the notes are also important. (Consider the contrast between the tune "Twinkle, Twinkle" and Haydn's famously surprising theme.) Hence the question arises: are those intervals represented in memory as such, or as patterns of pitch classes (musical scales)?

Dowling (1978) proposed that melodies are represented in memory by two components: contour and scale. The contour represents the pattern of ups and downs, along with relative rhythmic values. Since most melodic motion is in diatonic steps, stepwise motion is "unmarked" in the representation, whereas leaps are "marked" and their size in scale steps specified. When a melody is produced from memory, the contour is "hung" on the appropriate scale at the correct relative pitch level. Thus the overlearned scale pattern, which is common to a large number of melodies, consists of a fixed pattern of pitch classes, and thus provides the sizes of the intervals required to realize the melody.

This contour + scale approach contrasts with the theory that melodies are represented as a combination of contour and pitch intervals. For example, Trainor, McDonald, and Alain (2002) say: "Melodic information is thought to be encoded in two different 'relative pitch' forms, a domain-general contour code (up/down pattern of pitch changes) and music-specific interval code (exact pitch distances between notes)." And Trehub and Hannon (2006) say: "Adults' ability to recognize or reproduce familiar tunes necessarily depends on their encoding of finer pitch relations, specifically, intervals, or precise pitch distances between successive tones." These statements appear to suggest a model in which the beginning of "Twinkle, Twinkle" would be encoded in terms of contour and intervals in semitones as (0, +7, 0, +2, 0, -2). In contrast, Dowling's (1978) theory suggests that it would be encoded by having the contour select the pitches (*do, do, sol, sol, la, la, sol*) out of an internalized scale in a "moveable *do*" system (where *do* is assigned to the pitch of the tonic in whatever key the melody is transposed to). There is considerable converging evidence in favor of the latter approach (see Dowling, 1991; Dowling, Kwak, & Andrews, 1995):

(1) Recognition of two-tone intervals isolated from melodic context and transposed to novel keys is relatively poor (Cuddy & Cohen, 1976; Lee, Janata, Frost, Martinez, & Granger, 2015).

(2) In contrast, transposition of familiar melodic patterns (e.g., the NBC chimes) is quite accurate, even for nonmusicians (Attneave & Olson, 1971). This suggests that the melodies are more fundamental, and that the intervals are retrieved in terms of the melodies, rather than vice versa. (In other words, the melody gives us a means of tapping the knowledge stored in the tonal scale systems.)

(3) The tonal hierarchy (Krumhansl, 1990) operates in terms of pitch classes, not intervals.

(4) And, in a closely related point, the dynamic tendencies of tones are defined in terms of pitch classes in the tonal hierarchy. The seventh scale degree tends upward toward the tonic whether it is 1 semitone away (in the major mode) or 2 (in the natural minor or the Dorian modes). The second scale degree tends downward from 2 semitones above the tonic (major) or 1 semitone (Phrygian mode).

(5) Harmonic intervals such as thirds and sixths remain similar when inverted, showing again that it is the pitch classes preserved by inversion that are important, and not the intervals between them (Balzano & Liesch, 1982).

(6) Familiar melodies can be recognized even when the note-to-note interval pattern has been destroyed by inserting interleaved distractor notes between the notes

of the melody (Dowling, 1973; Dowling, Lung, & Herrbold, 1987), or by scrambling the notes of the melody into several octaves while preserving the pitch classes (Dowling, 1984; Idson & Massaro, 1978).

(7) Dowling (1986) showed that listeners with moderate levels of musical training automatically encode novel melodies they hear in terms of scale steps, not intervals. In contrast, nonmusicians tend toward interval encoding, and musical professionals can use either strategy depending on task demands.

One of the consequences of the theory that melodies are represented as combinations of contours and scales is that wrong notes in the melodies that violate the tonal scale pattern should be much more obvious and easy to identify than wrong notes that violate expected interval sizes. It has already been observed that out-of-key notes "pop out" of an otherwise uniform stream of pitches (Janata, Birk, Tillmann, & Bharucha, 2003). The experiment reported here is the latest in a series of studies in which we directly compare the rapidity with which listeners detect out-of-key wrong notes versus interval-distorting wrong notes (see Dowling, 2008; 2009; Raman & Dowling, 2015). The main difference between this study and our previous studies is that here we include a condition involving unfamiliar melodies, as well as familiar melodies. We expected that the tendency to rely on key membership as an index of whether a note is a wrong note would be even stronger with unfamiliar melodies, since there are no hard and fast rules governing interval sizes in melodies.

METHOD

Participants. Fifty-six students at the University of Texas at Dallas served in the experiment for partial course credit. Twenty-seven participants had less than 3 years of musical training, and 29 had 3 or more years and are characterized as moderately trained. Participants were assigned blindly to either the Familiar ($N = 21$) or the Unfamiliar ($N = 35$) condition.

Stimuli. The stimuli consisted of the first 16 to 24 notes of 32 familiar melodies that received the highest familiarity ratings in our previous studies (nursery tunes, folk tunes, patriotic songs, holiday songs, etc.), plus a stylistically similar group of unfamiliar folk songs drawn from Bronson (1976). These were all presented with their natural rhythms and at tempi that we judged to be comfortable for each melody. Each melody appeared twice in its respective session, and contained wrong notes in two different conditions. The wrong notes were introduced in a way that did not alter the contour of the melody. In the rare cases where the wrong note was part of a repeated pair of notes, the second note in the pair was also altered to match. There were eight kinds of wrong notes. The wrong notes were introduced by altering a target note either up or down from its original pitch, moving it 1 or 2 semitones, and landing on an in-key pitch or an out-of-key pitch. There were 64 trials in each session, in which these eight types of wrong note each occurred eight times. The wrong note could be introduced anywhere between the sixth note of the melody and the end. Previous research had shown that the up-down variable had negligible effects, and so we collapsed the data across that variable in the analyses reported here. The melodies were played by a MATLAB R2009b program as sequences of sine waves with linear on- and off-ramps to avoid clicks, and with 20-ms gaps between notes, and presented to listeners via high-quality headphones at comfortable levels. The program presented the stimuli in a different random order to each participant, and recorded their response times (RTs) to the wrong notes.

Procedure. Listeners were told that on each trial they would hear a melody that would often contain a wrong note, and that their task was to respond to the wrong note as quickly as possible. We told them to hold their fingers on the space bar and to press it as soon as they heard a wrong note. If they got to the end of the melody without hearing a wrong note, they were then to press the space bar to go on to the next trial. Each session lasted about 30 min.

RESULTS

We analyzed the data separately in the unfamiliar and familiar conditions, in each case scoring the responses for hits (correct detections of a wrong note in a window of 300 to 3000 ms following the onset of an actual wrong note), and for median RTs for correct detections (hits) for each condition (collapsed across direction, up vs. down) for each participant.

Unfamiliar Melodies. Table 1 shows the number of hits (out of 16) for each condition for listeners at the two levels of musical training. The hit rates are quite low, as might be expected for unfamiliar tunes. We ran an Analysis of Variance (ANOVA) with musical training as a between-groups variable, and key membership (in vs. out) and distance from original note (1 vs. 2 semitones) as within-group variables. To our surprise, key membership was not an important factor in detection of the wrong notes. Only distance from the original pitch was significant, $F(1,33) = 5.87$, $p = .02$, with alterations of 1 semitone ($M = 2.26$) detected more often than alterations of 2 semitones ($M = 1.66$).

Table 1. Number of hits with unfamiliar melodies (out of 16) by musically untrained participants and those with moderate amounts of training, for wrong notes that were in- or out-of-key, and 1 or 2 semitones removed from the original pitch. $N = 35$

Scale	Distance semitones	Untrained $N = 18$	Moderate $N = 17$	Means
IN	1	2.39	2.24	2.31
IN	2	1.39	1.65	1.51
OUT	1	1.83	2.59	2.20
OUT	2	2.06	1.53	1.80
	Means	1.92	2.00	

Table 2 shows the means of the median RTs for each condition. Because the hit rates in Table 1 were so low, only 16 of the original 35 participants produced hits in all four conditions. Those 16 split 8 and 8 on musical training. In the ANOVA on their data, only key membership was significant, $F(1,14) = 14.22$, $p = .002$, with RTs to out-of-key wrong notes

(*M* = 1039 ms) almost twice as fast as RTs to in-key wrong notes (*M* = 1960 ms).

When we analyzed the hit rates of those listeners with complete RT records, we found no significant effects of any of the variables. That is, they did not show the effect of interval size alteration found with the larger group of listeners.

Table 2. RTs (in ms) with unfamiliar melodies by musically untrained participants and those with moderate amounts of training, who achieved at least one hit in each of the four conditions, for wrong notes that were in- or out-of-key, and 1 or 2 semitones removed from the original pitch. *N* = 16

Scale	Distance semitones	Untrained *N* = 8	Moderate *N* = 8	Means
IN	1	2237	1764	2000
IN	2	1993	1847	1920
OUT	1	1260	651	955
OUT	2	1281	964	1122
	Means	1693	1306	

Familiar Melodies. Table 3 shows the number of hits for each condition for listeners at the two levels of musical training. In the ANOVA, the effect of training was significant, $F(1,19) = 12.88$, $p = .002$, with moderately trained listeners scoring more hits (*M* = 12.92) than untrained listeners (*M* = 9.50). Key had a strong effect, with out-of-key wrong notes detected more often (*M* = 12.62 out of 16) than in-key (*M* = 10.29), $F(1,19) = 29.33$, $p < .001$. Distance had an effect also, $F(1,19) = 14.94$, $p = .001$, with 2-semitone alterations (*M* = 12.12) more noticeable than 1-semitone alterations (*M* = 10.79). And distance interacted with experience, $F(1,19) = 11.86$, $p = .003$, such that distance was more important for untrained listeners (a 2.88 item gain for greater distance) than for the moderately trained (a 0.16 item gain).

Table 3. Number of hits with familiar melodies (out of 16) by musically untrained participants and those with moderate amounts of training, for wrong notes that were in- or out-of-key, and 1 or 2 semitones removed from the original pitch. *N* = 21

Scale	Distance semitones	Untrained *N* = 9	Moderate *N* = 12	Means
IN	1	6.67	12.00	9.71
IN	2	9.56	11.83	10.86
OUT	1	9.44	13.67	11.86
OUT	2	12.33	14.17	13.38
	Means	9.50	12.92	

Table 4 shows the RTs for each condition for both groups of listeners. Moderately trained listeners responded more quickly (*M* = 592 ms) than untrained listeners (*M* = 812 ms), $F(1,19) = 11.74$, $p = .003$. RTs were shorter to out-of-key wrong notes (657 ms vs. 716 ms), $F(1,19) = 5.92$, $p = .03$, and to 2-semitone alterations (*M* = 643 ms) than to 1-semitone alterations (*M* = 730 ms), $F(1,19) = 8.65$, $p = .008$. And this distance effect appeared mainly in the untrained listeners (901 vs. 724 ms) than in the moderately trained (603 vs. 582 ms), $F(1,19) = 5.42$, $p = .03$.

The key membership X distance interaction approached significance, $F(1,19) = 3.02$, $p = .10$, in which the effect of key membership was stronger with a 2-semitone alteration (*M* = 128 ms) than with a 1-semitone alteration (*M* = 46 ms).

Table 4. RTs (in ms) with familiar melodies by musically untrained participants and those with moderate amounts of training, for wrong notes that were in- or out-of-key, and 1 or 2 semitones removed from the original pitch. *N* = 21

Scale	Distance semitones	Untrained *N* = 9	Moderate *N* = 12	Means
IN	1	932	594	739
IN	2	797	614	693
OUT	1	869	611	721
OUT	2	651	550	593
	Means	812	592	

DISCUSSION

The most surprising result was that the listeners did not rely on key membership as a cue for wrong-note detection with the unfamiliar melodies. Distance of the wrong note from the original pitch had an effect, but it was the opposite of what was observed with the familiar melodies; that is, wrong notes 1 semitone away from the original were easier to detect than those 2 semitones away. We think that listeners were pursuing a strategy of noticing awkward sounding phrases, and taking those phrases as indicating the presence of wrong notes. Such a strategy risks producing RTs that are longer than even the 3-s window we provided. Hence it seems likely that many of the listeners who missed all the trials in one or another condition (and thus had to be deleted from the RT analysis) were following this strategy. When we looked at the detection data from the remaining listeners, it showed no effects of either distance or key membership.

These results complicate our understanding of the effects of experience on pitch encoding in memory for melodies. Dowling (1986), using an encoding specificity paradigm, demonstrated that moderately trained musicians automatically encode novel melodies they hear in terms of scale steps. However, if that is the case here, it seems that the moderately trained listeners are not developing a robust enough representation of the musical key of the unfamiliar melodies to use it in detecting wrong notes. Furthermore, it is clear from the present results that nonmusicians are not totally devoid of sensitivity to the scale steps (as they were in Dowling, 1986), since they show a strong effect of key membership on their RTs to the wrong notes in unfamiliar melodies.

Key membership is important in the detection of wrong notes in familiar melodies, leading to the detection of about 2.3 more wrong notes (out of 16) in the out-of-key conditions versus the in-key conditions. Distance was used as a cue, but with smaller effects, leading to an increase in detections of about 1.3 (out of 16). Note that the effect of distance was the opposite of that found with the unfamiliar melodies; here it was the wrong notes that were 2 semitones away from the original that were easier to detect. And distance was a more important cue for musically untrained listeners than for the moderately trained.

As with the unfamiliar melodies, RTs were strongly affected by key membership, with responses to out-of-key wrong notes about 59 ms faster. And there was a hint of an interaction between key membership and distance, such that both untrained and moderately trained listeners were especially quick to respond when an out-of-key wrong note was 2 semitones away from the original pitch. Distance affected RTs in general, but the effect was much more pronounced for the untrained listeners (about 177 ms faster for 2 semitones than for 1) than for the moderately trained (about 21 ms faster).

CONCLUSION

The present results replicate earlier studies in finding strong effects of key membership on the detection of and RTs to wrong notes in familiar melodies. We also found strong effects of the distance in semitones between the wrong note and the correct pitch. In unfamiliar melodies, the picture was more complicated, in that lack of key membership facilitated fast RTs, but not detection. Distance between the wrong note and the correct pitch in unfamiliar melodies seemed to function in a more global way than in familiar melodies, perhaps as by producing awkward sounding phrases that suggested the occurrence of wrong notes.

REFERENCES

Attneave, F., & Olson, R. K. (1971). Pitch as medium: A new approach to psychophysical scaling. *American Journal of Psychology, 84,* 147-166.

Balzano, G. J., & Liesch, B. W. (1982). The role of chroma and scalestep in the recognition of musical intervals out of context. *Psychomusicology, 2,* 331.

Bronson, B. H. (Ed.) (1976). *The Singing Tradition of Child's Popular Ballad*s. Princeton, NJ: Princeton University Press.

Dowling, W. J. (1973). The perception of interleaved melodies. *Cognitive Psychology, 5,* 322-337.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, *85,* 341-354.

Dowling, W. J. (1984). Musical experience and tonal scales in the recognition of octave-scrambled melodies. *Psychomusicology, 4,* 13-32.

Dowling, W. J. (1986). Context effects on melody recognition: Scale-step versus interval representations. *Music Perception*, *3,* 281-296.

Dowling, W. J. (1991). Pitch structure. In P. Howell, R. West, and I. Cross (Eds.) *Representing Musical Structure*. London: Academic Press, pp. 33-57.

Dowling, W. J. (2008, November). *Are melodies remembered as contour-plus-intervals or contour-plus-pitches?* Auditory Perception, Cognition, and Action Meetings (APCAM), Chicago, IL.

Dowling, W. J. (2009, November). *Directing auditory attention in terms of pitch, time, and tonal scale membership.* Poster session presented at the meeting of Psychonomic Society, Boston, MA.

Dowling, W. J., Kwak, S.-Y., & Andrews, M. W. (1995). The time course of recognition of novel melodies. *Perception & Psychophysics*, *57,* 197-210.

Dowling, W. J., Lung, K. M.-T., & Herrbold, S. (1987). Aiming attention in pitch and time in the perception of interleaved melodies. *Perception & Psychophysics*, *41,* 642-656.

Cuddy, L. L., & Cohen, A. J. (1976). Recognition of transposed melodic sequences. *Quarterly Journal of Experimental Psychology, 28,* 255-270.

Idson, W. L., & Massaro, D. W. (1978). A bidimensional model of pitch in the recognition of melodies. *Perception & Psychophysics*, *24,* 551-565.

Krumhansl, C. L. (1990). *Cognitive Foundations of Musical Pitch*. New York: Oxford University Press.

Lee, Y.-S., Janata, P., Frost, C., Martinez, Z., & Granger, R. (2015). Melody recognition revisited: Influences of melodic Gestalt on the encoding of relational pitch information. *Psychological Bulletin & Review*, *22,* 163-169.

Janata, P., Birk, J. L., Tillmann, B., & Bharucha, J. (2003). Online detection of tonal pop-out in modulating contexts. *Music Perception, 20,* 283-305.

Raman, R., & Dowling, W. J. (2015, August). *Effects of key membership and interval size in perceiving wrong notes: A cross-cultural study.* Paper session presented at the meeting of Society for Music Perception and Cognition (SMPC), Nashville, TN.

Trainor, L. J., McDonald, K. L., & Alain, C. (2002). Automatic and controlled processing of melodic contour and interval information measured by electrical brain activity. *Journal of Cognitive Neuroscience*, *14,* 430-442.

Trehub, S. E., & Hannon, E. E. (2006). Infant music perception: Domain-general or domain-specific mechanisms? *Cognition, 100,* 73-99.

Is pitch chroma discrimination automatic? — An EEG study

Tamar Regev[1], Israel Nelken[1,3] and Leon Y. Deouell[1,2]
[1]Edmond and Lily Safra Center for Brain Science, Hebrew University, Jerusalem, Israel
[2]Department of Psychology, Hebrew University, Jerusalem, Israel
[3]Department of Neurobiology, Hebrew University, Jerusalem, Israel

The perception of pitch is frequently described as helical, with a constantly increasing dimension of height, and a spherical dimension of pitch class (chroma). Thus, there is perceptual similarity between identical notes (e.g. C) played in different octaves. In this study we tested whether chromatic regularity - the same musical note dispersed across octaves - is detected automatically by the human brain. For this purpose, we used the mismatch negativity (MMN), an ERP component indexing automatic detection of deviation from auditory regularity. Musicians and non-musicians were trained for 10 minutes in classifying pure tones as C or G across 4 octaves, by singing the heard note in their preferred octave. Next they viewed a silent film while being passively exposed to two MMN protocols – 1) A protocol designed for testing chroma deviation – the standard (80% of stimuli) comprised of 4 different Cs, and the middle G (20%) served as deviant, and 2) A classic protocol of pitch deviation, comprised of 80% standard D and 20% deviant A. We found that perfect performers, who could perform the task effortlessly with almost no errors, differed from non-perfect performers by having significantly larger MMN amplitude in the classic protocol, indicating that automatic discrimination of pitch correlates with performance in the chroma discrimination task. The perfect performers had as well a larger P-300 response in the 'attend' block, in which they had to press a button each time that the chroma deviant appeared, compared to subjects with poorer performance. However, no chroma-related MMN was found in the passive block, even in the group of perfect performers. These results indicate that perceiving pitch chroma of pure tones is a task that can be learned and performed, but it is not implemented in the brain automatically and pre-attentively, and might require more elaborate cognitive processes.

The effect of familiarity on the time course of responses to modulation in classical music

W. Jay Dowling[1], Rachna Raman[1] and Barbara Tillmann[2]
[1]Behavioral and Brain Sciences, University of Texas at Dallas, USA
[2]Lyon Neuroscience Research Center (CNRL), France

We investigated listeners' responses to modulations in two Haydn string quartets from op.76 in D minor (no. 2) and C major (no. 3). Listeners at two levels of musical expertise (untrained, N = 60, vs. 6 or more years of training, N = 60) heard the 2-min exposition sections of the quartets, continuously rating how well the 12 possible probe tones fit the music, for 12 trials with each quartet. We assessed tonal hierarchy profiles for 11 5-s samples at approximately equal time intervals throughout the excerpt, with particular attention to points at which the music modulated, and correlated those profiles with the standard profiles of the keys in question. In general, more experienced listeners reflected the key changes in their ratings, whereas inexperienced listeners responded more globally, producing profiles that captured the principal keys of the excerpts as a whole. Here we compared ratings produced for the first three trials with a given excerpt, when it was relatively unfamiliar, with those for the last three trials, when they had heard it 9 or more times. As the more experienced listeners became more familiar with an excerpt, they tended to respond more globally, flattening out the shifts in their responses at points of modulation. In contrast, the inexperienced listeners became more attentive to the details with greater familiarity, registering the modulations more sharply. Both groups showed similar latencies of response to the modulations; in particular, in the C-major quartet they responded very slowly to the modulation from G major to G minor, and very quickly to the shift from G minor to Eb major.

Well Tempered Tuning Cannot Account for Low Purity Ratings of Isolated Voice Parts

Yke Schotanus
ICON, Universiteit Utrecht, The Netherlands
schotschrift.teksten@planet.nl

ABSTRACT

Because Western people usually hear melodies in well-tempered tunings, it is doubtful whether laymen would use a pure tuning as a reference when asked to judge whether a certain melody is sung in tune or not. In an earlier class room experiment among approximately 274 pupils (average age 15.6; SD 0.9) the isolated voice part of several songs was rated less in tune than the same voice part presented with the original accompaniment, although every mistuning was digitally corrected. It was hypothesized that these differences were due to the fact that music processing is more difficult without accompaniment defining rhythm and harmony. The aim of the current experiment was to test the competing hypothesis that purity ratings in conditions without accompaniment are affected by the fact that in a capella conditions a pure tuning is expected. To test this, 32 of the pupils that also participated in the earlier experiment (mean age 15.5; SD = 0.5) and 36 adults (mean age 56; SD = 15), rated the purity of the singing in six excerpts of the same songs in two conditions: pure tuning and well tempered tuning. The results indicate that in general there is no significant difference between purity ratings for excerpts in pure tuning, and excerpts in well tempered tuning. Among adults and pupils separately there are no differences. Among the total population, the pure tuning of one excerpt is significantly higher. Musical experience seems to have a small but interesting effect on purity ratings. Experienced musicians give relatively high purity ratings in both conditions, but tend to rate pure tunings slightly higher than well tempered ones, whereas listeners with moderate musical experience tend to rate well tempered tunings slightly higher.

I. INTRODUCTION

In a class room experiment among twelve fixed groups of pupils from the fourth grade of the highest levels of secondary education in the Netherlands (Schotanus, 2016 (ICMPC14)) the isolated voice part of two songs were rated significantly less in tune than the same voice part presented with the original accompaniment, although every mistuning was digitally corrected. Purity ratings for the isolated song parts of two other songs did not show significant differences with those for the accompanied versions, although, remarkably, they were even higher. Moreover, as table one shows, the isolated voice parts of a version of the songs in which each syllable of the lyrics was replaced by 'la' revealed a similar pattern of purity ratings. Furthermore, a series of Kruskal-Wallis tests showed that in most cases these results were the same across different groups.

The fact that purity ratings for isolated voice parts vary significantly across songs while they do not vary significantly for isolated voice parts of accompanied songs, might be interpreted as a support for the model for lyric processing I presented along with Musical Foregrounding Hypothesis (MFH)(Schotanus, 2015). In this model, I hypothesized that an accompaniment that explicates the rhythm and the implied harmony of a melody might ease the processing of the music in songs, at least in Western tonal music.

To assess this, the songs I chose for the earlier experiment had to be Western tonal songs with harmonic and rhythmic difficulties, such as out-of-key notes, modulations, and so called 'loud rests' (silences on the down-beat), components that are thought to complicate rhythm and harmony (Schotanus, 2015). Uncertainties about rhythm and tonality, caused by these elements, where thought to affect purity ratings in a negative way. In general, this turned out to be the case. The fact that, in a few cases, the purity ratings for the isolated voice parts where not negatively affected might (in line with the MFH) be explained by involvement. If either the lyric or the catchiness of the melody causes the listener to get involved in the music, the listener might overcome the difficulties in music processing and understand tonality, especially if the music is repeated several times. If purity can be understood as a specific kind of hedonic valence, this is in line with Berlyne (1989), among others, who states that there is an inverted-U shaped relationship between complexity and hedonic valence. This theory implies that an increase of complexity causes an increase in valence ratings if, and only if, the complexity has been overcome. If not an increase of complexity will cause a decrease in valence ratings.

Tabel 1: ***Purity ratings for isolated and accompanied voice parts by twelve groups of pupils (mean age 15)***

song	Isolated voice part 'la'		Accompanied song		Isolated voice part lyric	
	N	Mean (Median, SD)	*N*	Mean (Median, SD)	*N*	Mean (Median, SD)
1	72 (4)	2.52[b] (2; 1.26)	42 (2)	3.12 (3; .86)	64 (3)	3.40 (3; .93)
2	72 (3)	3.51 (4; .98)	49 (2)	3.24 (3; 1.03)	60 (3)	2.63[b] (3; 1.31)
3	46 (2)	2.69[b] (3; 1.15)	59 (3)	3.22[c] (3; 1.01)	58 (3)	3.31[c] (3; .94)
4	45 (2)	3.00[b] (3; .78)	62 (3)	3.56 (4; 1.01)	45 (2)	2.53[b, c] (3; .87)
All songs	235	2.90[d] (3; 1.15)	213	3.3 (3; 1,0)	227	3.00[d] (3; 1.1)

[a] Between brackets: number of groups
[b] A Mann-Whitney U test reveals: purity ratings are different from those for the accompanied version of the song
[c] A Kruskal-Wallis test reveals: purity ratings are different across groups; however, the deviant ratings do not really change the pattern
[d] A Kruskal-Wallis test reveals: purity ratings are different across songs

However, there might be an alternative explanation for the relatively low purity ratings for most isolated voice parts. Because these voice parts where recorded with piano accompaniment they are sung in well tempered tuning, while listeners might expect a pure tuning in a capella singing. I do not think this is a likely explanation because most of my pupils, living in a western culture, are used to hear melodies in a well tempered tuning. Moreover, it is known that non musicians without absolute pitch ability, hardly can tell whether a note is in or out of tune (Krumhansl, 1991) and listeners usually are more tolerant towards impurities of voices than of instruments (Hutchins, Roquet & Peretz, 2012). Furthermore, a single tone cannot be out of pitch, so it is likely that all purity ratings are based on the intervals in the music and even a person with absolute pitch ability is used to the intervals of well tempered tuning.

Nevertheless, the current experiment is created to examine whether listeners do use absolute pitches as a reference for purity ratings, instead of the pitches and intervals that belong to a well tempered tuning.

II. Method

Thirty-two pupils (mean age 15.3; SD 0,5) and thirty-six adults (mean age 54; SD 15) heard two textually and musically different sections of three different songs in two conditions (well tempered tuning and pure tuning) in a quasi-random order, each time divided by an atonal piano riddle.

A. Participants

Two groups of pupils that have participated in the earlier experiment where selected for this experiment too. Both are Dutch language and literature classes from the fourth grade of the highest level of secondary education in the Netherlands. One from SG Het Rhedens in Rozendaal, the other from Stedelijke Scholengemeenschap Nijmegen, in Nijmegen. The fact that these two groups are chosen and not others, is just a matter of formal consent, the choice is not based on their earlier performance.

The experiment was conducted after being reviewed and approved by the Ethics Review Board of the Utrecht Institute of Linguistics OTS (ETCL). Both school leaders have given signed permission, the parents of the children have been informed and have been given the chance to forbid their sons and daughters to participate and the pupils themselves where free to complete the questionnaires or not.

The adults are friends, acquaintances, colleagues and family members, who reacted to an e-mail containing the sound file and the questionnaire.

Every participant was told that the experiment was created to be sure of ‘something’ I might have found in the earlier experiment, but although they must have realized it had something to do with purity, I gave them absolutely no clue what the difference between the different versions might be.

B. Stimuli

The songs used in this experiment where the same as song 1, 2 and 4 of the previous experiment, two of which were rated significantly less pure in the condition without accompaniment but with lyrics. Of each song two fragments were chosen: the first verse or the first couple of verses, and the first occurrence of a second musical theme, either a different kind of verse, a bridge or a chorus. This might affect purity ratings because there is less repetition to help listeners to get used to the melodies, and processing second themes might be more difficult when they are not preceded by the first theme.

The original recordings are sung by the author, a male baritone and recorded by Christan Grotenbreg, a professional musician, in his studio. During the recording the singer was accompanied Christan Grotenbreg, who improvised the accompaniment based on te chord scheme served by the composer. He played a keyboard, connected to: ProTools 10 (Desktop recording). However, in this experiment only the isolated voice parts are used. The voice was recorded with a Neumann TLM 103 microphone, and an Avalon VT 737 SM amplifier. To avoid confounds concerning purity, voice-treatment software was used even for the original well tempered tuning: Waves Tune, Renaissance Vox compression and Oxford Eq. The (C-rooted) pure (or ‘absolute’) tuning was created digitally in Waves Tune, by Christan Grotenbreg, using the original (well tempered) recording.

The fragments were brought together in one 11 MB sound file in MP3 format by the author and a friend, Gerard Vuuuregge, using the computer software Audition. Within this file the excerpts were divided by silences and atonal piano riddles in order to prevent the tonality of a certain fragment to affect purity perception in the fragment following it. The song parts are presented in the following order, preventing two fragments of the same song to succeed each other and varying both song order and condition order:

- $4a_{\text{pure}}$,
- $1a_{\text{well tempered}}$,
- $2a_{\text{pure}}$,
- $1b_{\text{pure}}$,
- $2b_{\text{well tempered}}$,
- $4b_{\text{well tempered}}$,
- $2b_{\text{pure}}$,
- $4a_{\text{well tempered}}$,
- $1a_{\text{pure}}$,
- $2a_{\text{well tempered}}$,
- $1b_{\text{well tempered}}$,
- $4b_{\text{pure}}$.

C. Questions

For all fragments, listeners rated both purity of the singing and sureness about the own judgement on a five point likert scale. Furthermore they answered questions on gender, age, musical experience, musical background (Western or not), and hearing disabilities. Their names were not asked for. The questionnaires are strictly anonymous.

Concerning musical experience, I created a seven point scale. Points are adjusted with a maximum of seven, according to the following principles:

- plays an instrument (including voice) (1 – 3 years, 1 point; 4 – 6 years: 2 points; 7 – 9 years: 3 points, 10 years or more: 4 points);
- has received formal lessons in playing the instrument (1-3 years:1 point; 4 – 6 years: 2 points; more than 6 years: 3 points);

- has received formal lessons in music theory in school (0-2 years: 0 points (this is because most pupils in the fourth grade of Het Rhedens have received at least three years of formal music education at school); 3 years: 1 point; 4 years or more (or: has chosen music as part of the curriculum): 2 points);
- has received formal lessons in music theory apart from school (1 – 3 years: 1 point; 4 – 6 years: 2 points; more than 6 years: 3 points).

D. Statistical analyses

The significance of the differences between purity ratings for pure and well tempered versions of the same excerpts, and between sureness ratings for the same pairs of recordings, are tested using Wilcoxon signed rank tests in SPSS. To look whether there are individuals or groups with a clear bias for either pure or well tempered versions the sum of differences per individual was computed:

$$\sum_{k=1}^{n}\left(x_{k\,pure} - x_{k\,well\,tempered}\right).$$

Groups of individuals (adults versus pupils, or musicians versus non-musicians) where compared using Kruskal Wallis tests, Mann Witney U tests, or Ordinal Logistic Regressions using Generalized Logistic Modeling in SPSS. Furthermore, Repeated Measures Logistic Regressions were run on both purity and sureness ratings, using te Generalized Estimation Equation (GEE) in SPSS, in order to control for within subject effects, and to look whether there is a relationship between purity ratings and sureness ratings.

III. Results

In most cases there are no significant differences in purity ratings for pure or well tempered versions of the same isolated voice parts. In general for song 4a, 2a en 2b the well tempered tunings are rated more in tune, and for 1a and 1b the pure tunings. For 4b both tunings where rated almost exactly the same. However, as both a Repeated Measures Logistic regression (GEE), and a series of Wilcoxon signed rank tests reveal: these differences are not significant. Only one of these differences is: song $1a_{pure}$ is rated significantly more in tune than $1a_{well\ tempered}$ ($p_{Wilcoxon}$ = .025), but only when the purity ratings of the whole population are taken into account. Among adults or pupils only there are no significant differences. However, the difference between $2b_{well\ tempered}$ and $2b_{pure}$ is significant within one group of pupils. These outcomes suggest that there are differences in purity ratings across groups, and indeed a Kruskal-Wallis test shows that there are significant differences between purity ratings across groups for $2a_{well\ tempered}$ (p = .028), $2b_{pure}$ (p = .042) and $2b_{well\ tempered}$ (p = .048).

Tabel 2. Purity Ratings, Mean and SD, per Group, Song and Condition

	pupils	*adults*	*All*
$4a_{pure}$	3,67 (1,02)	3.66 (1.03)	3.68 (1.02)
$4a_{wt}$	3,64 (1,14)	4.01 (.87)	3.84 (1.02)
$1a_{pure}$	3,09 (.98)	3.22 (1.01)	3.16 (1.02)
$1a_{wt}$	2,85 (1.05)	2.91 (1.01)	2.91 (1,08)
$2a_{pure}$	2.82 (1.18)	3.4 (1.31)	3.14 (1.39)
$2a_{wt}$	2.82 (1.42)	3,52 (1.11)	3.16 (1,19)
$1b_{pure}$	3.24 (1.25)	3.23 (1.01)	3.25 (1,16)
$1b_{wt}$	2.94 (1.11)	3.11 (1.16)	3.15 (1.14)
$2\ b_{pure}$	2.58 (1.22)	3.01 (1.15)	2.85 (1.20)
$2b_{wt}$	2.76 (.87)	3.34 (1.16)	3.08 (1.06)
$4b_{pure}$	3.64 (1.08)	3.46 (1.07)	3.57 (1.08)
$4b_{wt}$	3.64 (.93)	3.46 (1.20)	3.55 (1.06)

In line with these results a Repeated Measures Logistic regression (GEE), shows that sureness ratings are significantly lower for both 1a (B = -.480, $Exp.(B)$ = .619; p = .005) and 2a (B = -.412, $Exp.(B)$ = .662; p = .024), and that the SD's for purity ratings are highest for 2a. Moreover, a Wilcoxon signed rank test reveals that the difference between sureness ratings between $2a_{pure}$ and $2a_{well\ tempered}$ is close to significance: p = .055. Incidentally, sureness ratings in general are higher than purity ratings.

Neither for purity ratings, nor for sureness ratings there are differences across age, music-cultural background and hearing disabilities. However, there are differences across gender and musical experience. As the GEE on purity ratings shows, women tend to give higher purity ratings than men, regardless of condition (B = .770, $Exp.(B)$ = 2.164; p .001), and experienced musicians give higher purity ratings than non-musicians ($B_{music=4}$ = .748, $Exp.(B)$ = 2.113; p = .019), as do highly experienced musicians (B_{music7} = 1.798, $Exp.(B)$ = 6.037; p = .001) (the groups with 5 en 6 points also give relatively high purity ratings, but probably the groups are too small to reach significance). On the contrary, purity ratings from listeners with moderate musical experience (music = 3), are lower than those from non-musicians, though not significantly lower. The same group (music = 3) also seems to be less sure about their purity ratings. As the GEE reveals, there sureness ratings are almost significantly lower than those from highly experienced musicians (p = .060).

An Kruskal Wallis test on a restructured version of the variable concerning musicianship, shows a similar pattern. Because some groups are rather small, as mentioned above, I restructured the variable on musical experience in four categories. Non-musicians (music 0 or 1), moderate experienced musicians (2 or 3), experienced musicians (4 and 5) and highly experienced musicians (6 and 7). As Figure 1. shows, listeners with moderate musical experience tend to rate well tempered tunings more in tune than pure tunings, and do so significantly more often than highly experienced musicians do (p = .034).

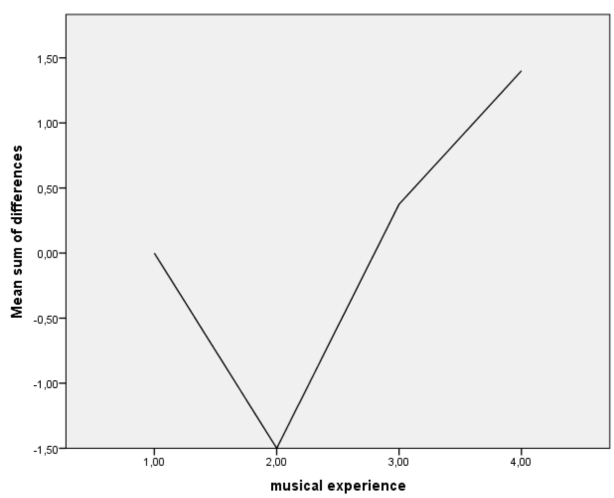


Figure 1. Mean sum of differences between purity ratings ($\Sigma(x_{k\,pure} - x_{k\,well\,tempered})$) from groups of listeners with a similar musical experience.

IV. Discussion

Listeners rated purity for six song parts in two conditions: well tempered tuning and pure tuning. Although in one case there is a significant difference between the purity ratings for the two versions of the same song part, it is clear that the relatively low purity ratings of isolated voice parts of songs recorded with piano accompaniment in the earlier experiment cannot be due to their well tempered tuning. The only excerpt for which the pure tuning is rated significantly higher than the well tempered one, excerpt 1a, is part of the song with the highest purity ratings for an isolated voice part in the initial experiment. They were even higher than those for $1a_{pure}$ in the current experiment. Moreover, $1a_{pure}$ is not rated especially high, but $1a_{well\ tempered}$ is rated relatively low. However, $1a_{well\ tempered}$ is the only excerpt with a significant difference between purity ratings from non musicians and experienced musicians. This excerpt seems to be very demanding for listeners with no or moderate musical experience. Furthermore, the difference between purity ratings for $1a_{pure}$ and $1a_{wel\ tempered}$ is much smaller than the differences between those for 2 or $4_{accompanied}$ and 2 or $4_{isolated\ voice\ part}$ in the original experiment.

Strange enough, the purity ratings for the well tempered versions of the excerpts of song 4 are much higher than those for the isolated voice part of the whole song in the first experiment. This might be due the fact that an important song section with a strongly deviant melody (the bridge) was not used for the current experiment.

Nevertheless, it is more likely to assume that the lower purity ratings for isolated voice parts in the initial experiment have something to do with the difficulty of processing unaccompanied voice parts. Possibly, this difficulty hampers purity perception more and more during the song, unless the melody or the lyrics can keep the listener involved, and repetition can habituate him to the difficulties of the melody.

An indication that purity ratings indeed are related to difficulties with the implied harmony caused by out-of-key notes and unusual intervals, is the fact that one adult listener made a remark about the unpredictability of the melodies, and another wrote that without a score it was difficult to get used to the semitones, not knowing where the flats and the sharps should be. Others remarked, more in general, that it is hard to tell what is pure or not, when there is no harmony to refer to. Therefore, one of them, a professional rock musician, wrote, he could not complete the questionnaire.

Several participants, both pupils and adults, reported that they heard something ‘Cher’-like unnatural in, or around my voice, a vocoder-effect, indicating that it was digitally fashioned. All of them said they preferred a more natural sound, be it less pure. As indeed there are digital modifications of my voice in both versions, but probably more in the pure ones, I asked two persons in which song parts the Cher-sound occurred. One reacted. According to him, the Cher-sound occurred in all songs, but it was salient in 4a and b_{pure}, and 2a and $b_{pure\ and\ well\ tempered}$. If he is right I must admit there is a correlation between purity ratings and Cher-sound effect, since purity ratings for the well tempered versions of 4a, 2a and 2b were higher than those for the pure versions of these song parts. However, it is strange that the Cher-sound effect would only affect purity ratings for the pure versions of 2a and b, not for the well tempered ones. Moreover, there is something else that song 2 and 4 have in common, and that is tonality. Both songs are in B flat, moving to D flat, B-flat minor, F and F minor, while song 1 is in E, moving toward A, B en F-sharp. Possibly, the scale tones in the B-flat-related scales are more different across tunings than those in the E-related scales.

However, the fact that there are significant differences between purity ratings for 2a and 2b across groups indicates that, perhaps, there might be other factors that play a part. Song 2 is the only song that is not a love song, but a song with a complex somewhat philosophical message, and on top of that, it is slower, contains more silences and implies many minor harmonies. Therefore, the music might be relatively demanding, while the lyrics might be less attractive for pupils than for adults, resulting in a loss of concentration among pupils.

Concerning the differences between musical experience, non-musicians, apparently, do not have a strong preference for either well tempered or pure tunings. However, musical training seems to make listeners less certain about their purity perception, resulting in lower purity ratings in general, with a preference for well tempered tuning. When musical experience increases listeners tend to become more self assured, resulting in higher purity ratings regardless of condition (in fact somehow all the excerpts are in tune), but with a small but increasing preference for pure tunings. This is remarkable because several musicians in this experiment (for example piano players and rock musicians) never use pure tunings.

V. CONCLUSION

The aim of this study was to investigate whether the relatively low purity ratings for isolated voice parts in an earlier class room experiment might be due to their well tempered tuning. The answer is: no. Neither the two groups of pupils who also participated in the earlier experiment, nor the 36 adults who were new to the materials, rated pure tunings of excerpts of the songs used in the earlier experiment significantly different from the original well tempered tunings of these fragments. The pure tuning of just one fragment was rated higher than the well tempered version of it, and only for the total population. Moreover, both pure and well tempered tunings received relatively low and inconsistent purity ratings, considering the fact that the purity of the recordings was digitally corrected. Therefore, the results of both experiments together seem to support the MFH-based idea that musical processing of isolated voice parts is difficult, resulting in lower purity ratings, expressing either a kind of tonal uncertainty or a general dislike for the song. Further research might reveal whether these lower purity ratings really are a matter of tonal uncertainty and therefore of purity perception, or whether they are in fact an expression of hedonic value.

Of course the implications of this experiment are limited. Perhaps, with another singer, or with songs in other tonalities, there would be more significant differences in purity ratings. However, that would not change the conclusions concerning the purity ratings for the isolated voice parts in the earlier experiment.

POSTSCRIPT

In the first weeks after this paper was submitted three new completed questionnaires arrived. After working up these subsequent reactions (new N_{adults} = 39, average age 56.6; SD 15.2) the difference between purity ratings for pure and well tempered versions for 2a is no longer significant (*p* = .078). However, among adults purity ratings for the well tempered version of 4a turned out to be significantly higher now than those for pure tunings. Both findings strengthen the conclusion that well tempered tunings cannot account for lower purity ratings of isolated voice parts.

The correlation between musicianship and purity ratings, became less convincing. A Kruskal-Wallis test for the whole pattern reveals that it is only close to significance (*p* = .059). However, a Mann-Witney test shows that the difference between groups 2 and 4 still is significant, (*p* = .008; with a Bonferroni correction *p* = .036).

Concerning tonality, an additional Repeated Measures Ordinal Logistic regression (GEE) revealed that purity ratings for well tempered tunings of songs in B_b (song 2 and 4) are significantly higher than those for well tempered tunings of the one song in E (song 1).

ACKNOWLEDGMENT

In the first place I would like to thank my participants for their collaboration; the parents of the ones under age and the school leaders of both RSG Het Rhedens and SSGN for their consent; and the Ethics Review Board of the Utrecht Institute of Linguistics OTS (ETCL) for reviewing my experimental design. Furthermore, I would like to thank my supervisors Prof. dr. E. Wennekes (Utrecht University), dr. F. Hakemulder (Utrecht University) and dr. R. Willems (Donders Institute, Nijmegen University), and statistics specialist Kirsten Schutter (UIL OTS, Utrecht University) for their support and comments. Finally, I thank the Dutch Government, NWO, and my employer, RSG Het Rhedens, for giving me the opportunity to avail myself of a PhD scholarship for teachers.

REFERENCES

Berlyne, D.E. (1971) *Aesthetics and Psychobiology*. New York: McGraw Hill.

Hutchins S., Roquet, C. & Peretz, I. (2012) The Vocal Generosity Effect: How Bad Can Your Singing Be? *Music Perception*, vol. 30, pp. 147–159. DOI: 10.1525/mp.2012.30.2.147

Krumhansl, C. L. (1991). Music psychology: Tonal structures in perception and memory. *Annual review of psychology*, *42*(1), 277-303.

Schotanus, Y.P. (2015) The Musical Foregrounding Hypothesis: How Music Influences the Perception of Sung Language. Ginsborg, J., Lamont, A., Philips, M. & Bramley, S. (Editors) *Proceedings of the Ninth Triennial Conference of the European Society for the Cognitive Sciences of Music*, 17-22 August 2015, Manchester, UK.

Schotanus, Y.P. (2016) Music Supports the Processing of Song Lyrics and Changes Their Contents: Effects of Melody, Silences and Accompaniment. *ICMPC14*. San Francisco, California

Visual rhythm perception of dance parallels auditory rhythm perception of music

Yi-Huang Su[1]
[1]Department of Movement Science, Technical University of Munich, Munich, Germany

Background
Musical rhythms are linked to human movements externally and internally. While this sensorimotor coupling underlies rhythm perception, it is unknown whether similar processes occur *visually* when observing music-related movements: dance. Given the action-perception coupling, watching rhythmic dance movements may elicit motor representations of visual rhythm.

Aims
I investigated visual rhythm perception of realistic dance movements with the following questions: Can observers extract different metrical levels from patterns of different body parts? Is visual tuning related to preferred tempo? Can one level serve as beat while another as subdivision? Are perceptual results supported by sensorimotor synchronization (SMS)?

Method
Participants observed a point-light figure performing Charleston and Balboa dance cyclically. In one task, they freely tapped to a tempo in the movement. In another, they watched the dance presented either naturally, or without the trunk bouncing, and reproduced the leg tempo. In two further SMS tasks, participants tapped to the bouncing trunk with or without the legs moving in the stimuli, or tapped to the leg movement with or without the trunk bouncing.

Results
Observers tuned to one level (legs) or the other (trunk) depending on the movement tempo, which was also associated with individual's preferred tempo. The same leg tempo was perceived to be slower with than without the trunk bouncing in parallel at twice the tempo. Tapping to the leg movement was stabilized by the trunk bouncing.

Conclusions
Observers extract multiple metrical periodicities in parallel from structured dance, in a similar way as for music. Lateral leg movements serve as beat and vertical trunk movements as subdivisions, as confirmed by tempo perception and the subdivision benefit in SMS. These results mirror previous auditory findings, and visual rhythm perception of dance may engage common action-based mechanisms as musical rhythm perception.

Musical rhythms induce long-lasting beat perception in listeners with and without musical experience

Karli Nave[1], Erin Hannon[1], Joel Snyder[1]
[1]Department of Psychology, University of Nevada Las Vegas, Las Vegas, Nevada, USA

Listeners are exposed to rhythmic stimuli on a daily basis, whether from observing others moving, listening to music, or listening to speech. Humans easily perceive a beat (quasi-isochronous pattern of prominent time points) while listening to musical rhythms, as evidenced by experiments measuring synchronized tapping or perceptual judgments. It is assumed that listeners infer the beat from regularly occurring events in the musical surface, but they sustain an internally driven metrical percept once the beat is inferred. Nevertheless, relatively few studies have attempted to disentangle the surface information from the internal metrical percept. We therefore attempted to measure the robustness of internally driven metrical percepts using a musical rich induction stimulus followed by a beat matching task with metrically ambiguous stimuli. During induction listeners heard an excerpt of unambiguous duple- or triple- meter piano music. They then heard a beat-ambiguous rhythm, which could be perceived as either duple or triple. In the probe phase, listeners indicated whether a drum accompaniment did or did not match the stimulus. Listeners readily matched the drum to the prior musical induction meter after the beat-ambiguous phase. Although musicians outperformed non-musicians, non-musicians were above chance. Experiment 2 examined the time course of the internal metrical percept by using the same task but varying the duration of the ambiguous phase. This revealed that listeners performed accurately and comparably for 0, 2, 4, or 8 measures of the ambiguous stimulus. Overall these results provide additional evidence for perception and long-lasting memory for musical beat.

Rhythmic skills in children, adolescents and adults who stutter

Simone Falk[1,3], Marie-Charlotte Cuartero[3], Thilo Müller[1,2], and Simone Dalla Bella[4,5,6,7]
[1]Department of German Philology, LMU, Munich, Germany
[2]Neurology (Stuttering Therapy), LVR Hospital, Bonn, Germany
[3]Laboratoire Parole et Langage, Aix-Marseille University, Aix-en-Provence, France
[4]EuroMov, University of Montpellier, Montpellier, France
[5]Institut Universitaire de France, Montpellier, France
[6]BRAMS, Montreal, Canada
[7]WSFiZ, Warsaw, Poland

The production of fluent speech is related to the capacity to precisely integrate auditory and motor information in time. Individuals who stutter have shown temporal deficits in speech production which are possibly linked to altered connectivity and functioning of the neural circuits involved in speech timing and sensorimotor integration. The aim of this study is to examine if individuals who stutter show deficits which extend to non-verbal sensorimotor timing as well. In a first experiment, twenty German-speaking children and adolescents who stutter and 43 age-matched controls were tested on their sensorimotor synchronization skills. They were asked to tap with their finger to rhythmic auditory stimuli such as a metronome or music. Differences in sensorimotor timing between participants who stutter and those who do not stutter were found. Participants who stutter showed poorer synchronization skills (i.e., lower accuracy or consistency) than age-matched peers. Adolescents who stutter were more impaired than children, particularly in consistency. Low accuracy resulted from the fact that participants who stutter tapped earlier in relation to the pacing stimulus as compared to controls. Low synchronization consistency (i.e., higher variability) was observed in particular in participants with severe stuttering. In a second experiment, we tested 18 French-speaking adults who stutter and 18 matched controls using the same tasks. In addition, to assess whether stuttering was accompanied by a beat perception deficit, the participants were tested on a beat alignment task. Adults' synchronization skills were similar to those of children and adolescents (i.e., more variable and less accurate than controls). In contrast, the two groups performed similarly well in the perceptual beat alignment task. In sum, timing deficits in stuttering are not restricted to the verbal domain and are likely to particularly affect processes involved in the integration of perception and action.

Assessment of rhythm abilities on a tablet-based mobile platform

Theodore P. Zanto[1], Namita Padgaonkar[1], Adam Gazzaley[1,2]
[1]Department of Neurology, University of California San Francisco, USA
[2]Departments of Physiology & Psychiatry, University of California San Francisco, USA

In our digitally enriched environment, there is now the potential to use consumer grade mobile technologies to collect research data, which could enable remote data collection from participants at home. As a first step to assess the utility of a tablet-based mobile platform in characterizing rhythm abilities, a custom rhythm game was developed in Unity and played by participants in-lab on a Windows Surface 3 tablet. The rhythm game has 27 levels that consist of combinations of 3 tasks (tap with the metronome, tap between metronome, call-and-response to metronome), 3 tempos (350 ms, 525 ms, 750 ms inter-onset-interval), and 3 stimulus types (audio, visual, and audio-visual). Four groups of participants were recruited based on age (younger (18-35 years), older (60-80 years) adults) and musical experience (non-musician (< 5 years musical training), musician (> 10 years musical training)). Rhythm abilities were assessed based on accuracy, precision and variance. Results replicate previously observed findings in that musicians outperformed non-musicians, younger adults outperformed older adults, fast tempos were most difficult, and audio-visual stimuli as well as on-beat synchronization were easiest for all participants. A week later all participants played the rhythm game a second time and test-retest reliability was good to excellent (intra-class correlation coefficients ranged from 0.72 to 0.83). Results were compared to other assessments of perceptual, motor and cognitive functions known to decline in aging. Across all participants, better rhythm abilities significantly correlated with faster response times, higher working memory capacity, better visuospatial selective attention, and lowered multitasking costs. Together, these results indicate tablet-based mobile platforms may serve as a useful tool for remote data collection and are sensitive enough to discriminate rhythm abilities based on age and musical experience.

A Bayesian approach to studying the mental representation of musical rhythm using the method of serial reproduction

Elisa Kim Fromboluti[1], Patrycja Zdziarska[2], Qi Yang[3] and J. Devin McAuley[1,4]
[1]Psychology, College of Social Science, Michigan State University, East Lansing, MI, USA
[2]Plant Biology, College of Natural Science, Michigan State University, East Lansing, MI, USA
[3]GN ReSound, Chicago, IL, USA
[4]Neuroscience, College of Natural Science, Michigan State University, East Lansing, MI, USA

Adult listeners lose some ability to recognize and categorize rhythmic patterns present in infancy through enculturation to a musical tradition. Some constraints on rhythm perception, however, are likely to be innate. The present research investigated the mental representation of musical rhythm using the method of serial reproduction. Participants were presented with brief sequences, and then reproduced what they heard or saw by tapping. The reproduced rhythm served as the target rhythm for the next participant. Data were collected in a public setting using a custom-designed app at ArtPrize, a 19-day art competition in Michigan. Over 400 individuals varying in age and musical experience participated. Modality, rhythmic structure, and tempo of initial sequences were manipulated. A Bayesian analysis is being conducted to examine how each link in the chain of reproductions systematically distorts the information passed from person to person. We hypothesize that (1) individual priors for tempo center on previously observed preferred tempo within a range of inter-onset-intervals (IOIs) between 400–600-ms; and (2) priors for meter may depend on both cultural expectations and inherent constraints. Consistent with the prediction that regardless of the rhythm initiating a chain, iterations of reproductions will converge to culturally- or cognitively-predicted preferred tempo or meter, preliminary analysis of the tempo chains reveals systematic drift in the reproduced IOIs to the predicted preferred tempo range for both auditory and visual sequences. Analysis of chains involving the reproduction of simple and complex meters is in progress. To our knowledge, this approach to studying the influence of biases on rhythm perception is unique: the method of data collection enables us to engage a broader population of non-experts in music cognition than is typically seen in the lab using a method that has not yet been applied to examine the mental representation of rhythm.

Absolute versus Relative Judgments of Musical Tempo: The Roles of Auditory and Sensorimotor Cues

Justin London[1], Marc Thompson[2], Birgitta Burger[2] and Petri Toiviainen[2]
[1]Department of Music, Carleton College, Northfield, MN, USA
[2]Department of Music, University of Jyväskylä, Jyväskylä, Finland

Background
Listeners have good memories for the absolute tempo of well-known songs and are sensitive to relative tempo differences in simple rhythms, but in a recent study (London, et al, *Acta Psychologica* 2016) participants made systematic errors when presented with time-stretched vs. original versions of popular songs: sped-up slow songs were rated faster than slowed down fast songs, even though their absolute tempos (measured in beats per minute) were the same.

Aims
To study these errors and their possible sources, two follow-up experiments were conducted. The first varied BPM rate, low frequency spectral flux, and dynamics, and the second involved synchronized self-motion (tapping).

Method
Stimuli from the previous study were re-used in both experiments, 6 classic R&B songs (matched high/low flux pairs at 105, 115, and 130 BPM), plus versions time-stretched ±5%; participants rated the speed (not beat rate) on a 7-point scale. In Exp1 stimuli were presented at two loudness levels (±7db); in Exp2 stimuli were presented at a single dynamic level and participants tapped along to the beat.

Results
In Exp1 significant MEs were found for BPM, Flux, and Loudness, and a significant BPM x Flux interaction. High flux songs were rated faster than low flux, louder faster than softer, and flux and loudness had greater effects at slower tempos. Overall, the confusion between absolute vs. relative tempo found in the 2016 experiment again occurred. Pilot results from Exp. 2 found that while tapping ameliorated the relative vs. absolute tempo errors somewhat, it also led to octave errors at faster BPM levels, and correspondingly slower tempo judgments.

Conclusions
Collectively, these results support an account of tempo in which "faster" is equated with either relative or absolute increases in observed and/or expended energy.

The perception of a dotted rhythm embedded in a two-four-time framework

Chinami Onishi[1], Yoshitaka Nakajima[2], and Emi Hasuo[3]
[1]Kyushu University, Fukuoka, Japan
[2]Department of Human Science/ Research Center for Applied Perceptual Science, Kyushu University, Fukuoka, Japan
[3]College of Education, Psychology and Human Studies, Aoyama Gakuin University, Tokyo, Japan

Previous studies had shown that there are systematic discrepancies between played rhythms that are physically correct and that are subjectively correct. The purpose of the present study was to examine how a beat framework would affect such a discrepancy observed in a dotted rhythm. A dotted rhythm consisting of a dotted quaver and a semiquaver embedded in two-four time was presented to 13 participants. Very short pure-tone bursts (1000 Hz) were used to mark the onsets of the notes; the whole pattern sounded as if played on a percussion instrument. The location of the dotted rhythm was varied between the first beat of the second bar and the second beat of the third bar. The tempo was 80, 100, or 125 bpm. A music notation to be referred to was also presented. The participants were instructed to adjust the physical ratio between adjacent durations so as to make them subjectively correct. The results, by and large, indicated a tendency that the adjusted physical proportion of the dotted quaver was larger than the notated proportion, either in the order of the dotted quaver and the semiquaver (notating 3:1) or in the opposite order (1:3). This tendency was larger when the semiquaver preceded the dotted quaver, except when the dotted rhythm was located on the first beat of the second bar. No effect of the location within the bar (whether the dotted rhythm was on the first or on the second beat) was observed. The physical proportion of the dotted quaver (that was subjectively correct) tended to be larger than the notated proportion, which was in line with previous studies. The results also suggested that the location of a dotted rhythm might influence the perception of the duration ratio.

It’s open to interpretation: Exploring the relationship between rhythmic structure and tempo agreement in Bach’s *Well Tempered Clavier*

Michael Schutz[1,2], Raven Hebert-Lee[1], Anna Siminoski[2]
[1]School of the Arts, McMaster University, Canada
[2]Department of Psychology, Neuroscience and Behaviour, McMaster University, Canada

Why do some compositions compel performance at certain tempi whereas others “sound right” under a variety of interpretations? Although J.S. Bach’s *Well Tempered Clavier* has been reprinted numerous times and recorded extensively, his “indented tempi” remain elusive. In his liner notes András Schiff (2011) openly confesses to wondering “what is the right tempo…and how do we find it?”, and Wanda Landowska (1954) acknowledges that some of her tempos “will probably be surprising.” Renowned artists have disagreed strikingly on interpretation—Glenn Gould recorded the E minor Fugue at twice the tempo of Rosalyn Tureck, and Anthony Newman the B minor Prelude at three times that of Fredric Gulda! However this discrepancy is not universal – performances of the Eb minor Prelude typically similar. Willard Palmer’s survey (used previously by Poon & Schutz, 2015 to quantify timing), affords a novel opportunity to explore this intriguing issue.

Palmer’s landmark analysis covers 20 authoritative interpretations –13 albums, 5 notated editions, and 2 commentaries. We defined tempo agreement using the standard deviation of tempo choices, and calculated rhythmic variability using the normalized Pair-wise Variability Index (nPVI). Our analysis revealed that pieces with high rhythmic variability led to strong tempo agreement and pieces with low rhythmic variability (i.e. isochronous) led to low tempo agreement. For example, the Eb minor Prelude exhibited the greatest rhythmic variability, as well as the strongest tempo agreement. In contrast both the B minor Prelude and E minor Fugue consist of isochronous rhythmic structures, and are consequently performed a wide range of tempi. The relationship between rhythmic structure (nPVI) and tempo agreement raises intriguing questions regarding the interaction between composers’ structural choices and performers’ interpretive decisions. An interactive visualization is now available online at www.maplelab.net/bachTempi.

Beat deaf persons entrain: Evidence for neural entrainment despite impaired sensorimotor synchronization

Benjamin G. Schultz[1], Pauline Tranchant[2], Baptiste Chemin[2,3], Sylvie Nozaradan[2,4], and Isabelle Peretz[2]
[1]Department of Neuropsychology & Psychopharmacology, Faculty of Psychology & Neuroscience, Maastricht University, The Netherlands
[2]International Laboratory for Brain, Music & Sound Research, Université de Montréal, Québec, Canada
[3]NOCIONS Laboratory, Institute of Neuroscience, Université catholique de Louvain, Brussels, Belgium
[4]MARCS Institute, Western Sydney University, New South Wales, Australia

There is evidence that healthy individuals entrain to the beat via neural oscillations that synchronize with, and adapt to, regularities within an auditory sequence. Here we examined whether beat deaf participants, who have experienced lifelong difficulties to synchronize movements to the musical beat, would display abnormal neural entrainment when listening and tapping to two complex rhythms that were either strongly or weakly beat inducing, at two different tempi (5Hz and 2.67Hz base frequencies). First, participants were instructed to detect tempo accelerations while listening to the repeating rhythm. In a second phase, they participants were instructed to tap the beat of the rhythm. Time-frequency analyses identified the frequencies at which neural oscillations occurred and the frequencies at which participants tapped. We hypothesized that behavioral beat entrainment (i.e., tapping) is impaired for beat deaf persons compared to controls while neural entrainment (i.e., steady-state evoked potentials) is comparable for beat deaf persons and controls. Results indicate that, for both rhythms and tempi, beat deaf participants did not demonstrate a reduction of entrainment compared to controls at the relevant beat frequencies during listening; Bayes factor *t*-tests revealed considerable evidence in favor of the null hypothesis indicating that entrainment by beat deaf demonstrated no deficit compared with controls. Conversely, beat deaf participants demonstrated a reduced ability to synchronize with the beat at the frequencies of interest compared with controls when tapping. Findings suggest that beat deafness is primarily a motor/sensorimotor deficit and not a perceptual disorder as implied by the term "beat deafness."

Beat Keeping in a Sea Lion as Coupled Oscillation

Andrew Rouse[1], Peter Cook[2], Edward Large[3], and Colleen Reichmuth[1]
[1]Institute of Marine Sciences, University of California Santa Cruz, Santa Cruz, California, USA
[2]Department of Psychology, Emory University, Atlanta, Georgia, USA
[3]Department of Psychological Sciences, University of Connecticut, Storrs, Connecticut, USA

The capacity for entraining movement to external rhythms — beat keeping — is ubiquitous in humans, but its evolutionary history and neural underpinnings remain a mystery. Recent findings of beat keeping in non-human animals pave the way for a novel comparative approach to assessing the origins and mechanisms of rhythmic behavior. The most reliable non-human beat keeper to date is a California sea lion, Ronan, who was trained to match head movements to isochronous repeating stimuli and subsequently showed spontaneous generalization of the ability to novel tempos and to the complex rhythms of music. Does Ronan's performance rely on the same neural mechanisms as human rhythmic behavior? Here we present findings from a recent study where we presented Ronan with simple rhythmic stimuli at novel tempos. On some trials, we altered either tempo or phase in the middle of a presentation. Ronan quickly adjusted her beat-keeping behavior following all perturbations, recovering her consistent phase and tempo relationships to the stimulus within a few beats. Ronan's performance was consistent with predictions of mathematical models describing coupled oscillation. Specifically, a model relying solely on phase coupling strongly matched her behavior, and the model was further improved with the addition of period coupling. These findings show that Ronan's behavior is similar in most respects to human beat perception and synchronization, and suggest that human rhythmic faculty may be rooted in evolutionarily conserved mechanisms.

Rhythmic Perceptual Priors Revealed by Iterated Reproduction

Nori Jacoby[1], Josh McDermott[1]
[1] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

Western musical notation is built on temporal intervals related by simple integer ratios. However, the reason for the importance of these ratios is unclear, as some traditions feature non-integer rhythmic patterns (Polak & London 2014). The issue remains unresolved in part because prior studies have either relied on explicit judgments of experts (Desain & Honing 2003; Clarke 1984), or have predominantly tested Western musicians (Povel 1982;Repp et al. 2011).

We introduce a novel procedure adapted from computational concept modeling (Xu et al., 2010) in which a listener's reproduction of a 2- or 3- interval rhythm is used to infer the internal "prior" probability distribution constraining human perception. Listeners are presented with a random seed (interval ratios drawn from a uniform distribution), and reproduce it by tapping or vocalizing. The reproduction is then substituted for the seed, and the process is iterated. After a few iterations the reproductions become dominated by the listener's internal biases, and converge to their prior.

When tested in this way, Western musicians consistently converged to attractors near integer-ratios. Integer-ratio biases were evident even in non-musicians. To explore whether these biases are due to passive exposure to Western music, we tested members of the Tsimane', an Amazonian hunter/gatherer society. Despite profoundly different cultural exposure, the Tsimane' also exhibited a prior favoring integer ratios, though the prominence of certain ratios differed from Westerners.

Our results provide evidence for the universality of integer ratios as perceptual priors. However, differences exist between groups in the relative strength of rhythmic perceptual categories, suggesting influences of musical experience. Both findings were enabled by our method, which provides high-resolution measurement of rhythmic priors regardless of the listener's musical expertise.

Ensemble synchronization and leadership in Jembe drumming music from Mali

Nori Jacoby[1,*], Justin London[2,*], Rainer Polak[3,*]
[1]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, USA
[2]Department of Music, Carleton College, USA
[3]Institute for World Music, Cologne University of Music and Dance, Germany
[*]all authors contributed equally to this work

Background
Recent studies of ensemble synchronization provide valuable insight into the question of musical leadership, but thus far these methods have been only applied to Western music (Keller 2013; Wing et al. 2014). Performers in a Jembe ensemble from Mali (typically 2-4 musicians) inhabit distinct musical roles: lead drum, repertoire-specific timeline, and one or two ostinato accompaniment parts, making this repertoire ideally suited to investigate ensemble entrainment.

Aims
We evaluate two different computational models of ensemble synchronization to study implicit leadership roles as expressed in timing relationships between performers, and expand the application of this methodology in a cross-cultural comparative perspective.

Method
We analyzed 15 performances of three pieces of Jembe percussion music recorded in Bamako, Mali in 2006/07. A semiautomatic onset detection scheme yielded a data set of ~40,000 onsets with a temporal resolution of ±1ms. Two different statistical approaches, linear phase correction modeling (Vorberg and Schulze 2002) and Grainger causality (Barnett and Seth 2014), were used to analyze the timing influences between the parts.

Results
We found that whereas no single part is a clear leader, there is one consistent follower: the lead drum, which adapts to the timeline and accompaniment parts to a far greater extent than vice-versa. This is significantly different than the timing behaviors found for Western string quartets (Timmers et al 2014; Wing et al. 2014).

Conclusion
Our finding that the lead drum has the most adaptive role in terms of ensemble timing suggests a reconsideration of the concept of leadership in Jembe ensembles, and perhaps more generally. These findings also complicate the Africanist concept of timeline as the central timing reference. Rather, the accompaniment plus timeline parts together play a time-keeping role in the jembe ensemble.

Groove in context: Effects of meter and instrumentation

Richard Ashley[1]
[1]School of Music, Northwestern University, Evanston, IL, USA

Background
Groove—that aspect of a musical stimulus that elicits a tendency to entrain or move along with it—is of empirical and theoretical interest (Danielsen 2006, Janata, Tomic, & Haberman 2012, Madison et al 2011, Witek et al 2014). Prior work has focused on groove and genre, microtiming, or tempo, with less attention to specifics of musical structure.

Aims
This project investigates groove as a function of multi-voice rhythmic structure. Our emphasis is on the role of interactions of accent structures in different musical lines.

Methods
Stimuli are taken from two different sources: rhythm section recordings (drums, bass, and other instruments, e.g. guitar and keyboards) and full recordings (excerpts with vocals and instrumentation). Stimuli vary by number and combinations of lines (one, some, or all rhythm instruments, or song segments including only rhythm section instruments or others, such as horns), and also by selected transformations of order (such as retrograde of original rhythms). Participants give ratings for groove and also move to the music as desired, not limited to manual tapping.

Results
Data are still being collected so results are preliminary. Rhythmic structure has a significant influence on groove, more with regard to entrainment than to global ratings of groove. Listeners seek 'stable' patterns of entrainment to rhythmic schemata in popular music and entrain to these, which are not necessarily predicted by notation or instrumental conventions (i.e. bass drum = downbeats), although rhythm section instruments are more influential than others. High-groove rhythms tend toward 'front loading,' with stronger and more predictable periodic structure at the beginning of metric units and more variability as the metric unit progresses.

Conclusions
A sense of groove emerges from the interactions of musical lines within metric frameworks, with listeners responding to both stability (predictability) and diversity (novelty) of rhythmic structure.

Single gesture cues for interpersonal synchronization

Esther Coorevits[1], Dirk Moelants[1] and Marc Leman[1]
[1]IPEM Dept. of Musicology, Ghent University, Ghent, Belgium

Synchronization is an essential aspect of ensemble performance. Musicians use different strategies to start playing together and keep the tempo. One of the methods commonly used is sniffing. Intriguingly, the sound of a single sniff does not contain explicit intervallic information. Still, a sniff can make musicians synchronize and communicate tempo successfully. Similarly, conductors often use a single gesture to set the tempo of the music. The aim of this study was to discover which features of these single gestures make the communication of timing possible. We hypothesize that the sniff functions as a sonification of a visual cue, translating the gesture of a conductor in an auditory trace. To test this, two experiments were conducted. In the first experiment, 2x20 musicians were asked to synchronize single claps and clap simple rhythms in 3 tempi and 3 meters, only using a single sniff. In the second experiment, 10 conductors were asked to conduct an ensemble of two musicians clapping the same rhythms as in the first experiment. The conductors were asked to indicate the tempo and the meter of these rhythms prior to the first note with only one single gesture. The results show that the maximum intensity of the sniff is significantly higher in the fast tempo, while the duration becomes significantly shorter. Similarly, the peak acceleration in the conductors' gestures increases with tempo while the duration of the gesture becomes shorter. By means of functional data analysis, closer investigation of both sniff intensity and conductors' velocity and acceleration patterns shows an overall similarity in shape. No clear influence of meter was found. These results show that single gestures can communicate aspects of timing through their shape. This supports the idea that the communication of timing does not has to rely on event-based aspects only, but that emergent, continuous aspects of time are also sufficient for ensemble synchronization.

Perceived difference between straight and swinging drum rhythms

Andrew V. Frane[1]
[1]Department of Psychology, UCLA, Los Angeles, CA, USA

Swing is a popular rhythmic technique that systematically delays the onset of each even-numbered division of a bar (at a particular metric level, e.g., 16th-note swing delays each even-numbered 16th-note subdivision). This creates a "swinging" pulse of alternating long and short divisions, as opposed to a "straight" pulse of isochronous divisions. The present study used two experiments to explore college students' thresholds for perceiving a difference between straight rhythm and swinging rhythm. The first experiment used a signal detection paradigm: Participants listened to pairs of looped drum patterns and judged whether the two patterns sounded "identical" or "different" in timing; in each pair, the patterns either were either exactly alike or differed only in that one pattern was straight and the other was swinging. Two independent variables were manipulated from trial to trial: magnitude of swing (when present), and rhythmic density. The second experiment used the method of adjustment: Participants listened to continually-looped drum patterns that were straight when a button was depressed and swinging when the button was released, and freely adjusted the swing's magnitude (the "swing ratio") until the point of just-noticeable-difference was determined. Two independent variables were manipulated from trial to trial: tempo, and the loudness discrepancy between odd-numbered and even-numbered 16th-note divisions. Results of the two experiments are discussed in terms of how the manipulated variables affected the threshold for detecting a difference between straight and swinging versions of a rhythm.

Predicting temporal attention in music with a damped oscillator model

Brian K. Hurley[1,2], Lauren K. Fink[2,3], Petr Janata[1,2,3]
[1]Department of Psychology, University of California Davis, Davis, CA, USA
[2]Center for Mind and Brain, University of California Davis, Davis, CA, USA
[3]Neuroscience Graduate Group, University of California Davis, Davis, CA, USA

Convergent evidence from psychophysics, neuroimaging, and neurophysiology suggests that the rhythmic structure of auditory stimuli can dynamically enhance temporal processing. A damped oscillator model that uses reson filters to estimate a stimulus' metric structure (Tomic & Janata, 2008) holds promise in predicting listeners' temporally graded expectancy. However, this possibility remains to be tested. In the work presented here, we aim to psychophysically map the dynamic allocation of attention in musical auditory scenes. Moreover, we use these behavioral data to examine the precision with which the oscillator model predicts temporal attention in musical scenes. We used an adaptive threshold procedure to probe temporal attention. In the first experiment, participants (N = 27) detected transient (200 ms) intensity increment deviants embedded in continuously repeating multi-timbral percussion loops. Critically, deviant time points were selected on the basis of output from the oscillator model, and separate thresholds were estimated for each probed position. Results indicate that Tomic and Janata's oscillator model significantly predicts deviant detection performance, such that moments of higher modeled salience correspond to lower detection thresholds, i.e. better performance. However, given that intensity increases are perceptually analogous to intensity accents, it is possible that transient intensity increments may perturb one's perception of metric structure. As such, in a second experiment, we altered the procedures such that participants detect intensity decrements. Preliminary results (N=12) replicate the results of Experiment 1, such that higher modeled salience significantly predicts lower absolute-value thresholds for decrement detection. Taken together, results from these experiments suggest the utility of Tomic and Janata's resonator model in predicting temporal attending behavior in complex musical scenes.

Rhythmic Complexity in Rap

Adam Waller[1] and David Temperley[1]
[1]Eastman School of Music, Rochester, NY, USA

It is generally agreed that rap songs vary in their rhythmic complexity; it has also been claimed that rap has gotten more rhythmically complex over time. As yet, however, few attempts have been made to quantify the rhythmic complexity of rap or to measure its historical development.

We hypothesize that the perceived rhythmic complexity of a rap depends significantly on the alignment between the syllables of the rap and the underlying meter. A rap with many syllables (especially stressed syllables) on weak beats should seem less compatible with the meter, and therefore more complex. In this paper we test this hypothesis with an experiment and a computational model.

For the experiment, synthesized versions of actual rap songs were created consisting only of vocal syllables of two types, stressed (e.g. "*TAH*") and unstressed (e.g. "tuh"), plus a drum track. Subjects had to judge how well the synthesized rap "fit" the underlying meter conveyed by the drum track.

A generative probabilistic model was devised to predict listeners' judgments. The model assigns each rap a probability, based on the type of syllable event (stressed, unstressed, or none) occurring at each beat, distinguishing between beats depending on their position in the measure. (A corpus of 20 rap songs was used to set the model's parameters.) The model shows a significant correlation with human judgments of "fit" between the rap and the drumbeat. The model is also tested on the probability it assigns to the corpus data itself. The role of metrical position and syllabic stress in the model's performance will be examined.

Several further issues will be considered: The change over historical time in the complexity of rap, the role of "intra-opus" probabilities (i.e. repeated rhythmic patterns within a piece) in rhythmic complexity, and the role of anticipatory syncopation (the tendency for accented events to fall just before a strong beat).

Exploring rhythmic synchronization abilities in musicians using training-specific movements

Anna Siminoski[1], Fiona Manning[1] and Michael Schutz[2,1]
[1]Department of Psychology, Neuroscience and Behaviour, McMaster University, Hamilton, Ontario, Canada
[2]The School of the Arts, McMaster University, Hamilton, Ontario, Canada

Musicians are more accurate than nonmusicians at rhythmic synchronization tasks. Many studies use finger tapping as a means to investigate timing abilities, however it is unclear the extent to which motor training on musicians' respective musical instruments extend to finger tapping abilities. Here we examined how musicians' training-specific movements compare in both consistency and accuracy to finger tapping in a synchronization task. Participants of different musician groups listened to an isochronous sequence of beats and identified the final beat as being either consistent or inconsistent with the preceding sequence. Participants either synchronized their movements to the sequence (movement conditions) by using a drumstick, tapping with a finger, or playing their instrument, or they listened without moving (no-movement condition). Previously we found that musicians were significantly better at identifying the timing of the final beat when allowed to move than when not moving. In the present study we examined the synchronization data for each motor effector in the musical groups by assessing the degree of variability between movements and the asynchronies between the movements and the beat sequence. We observed that stick tapping is less variable and more synchronous for all musicians and nonmusicians compared to finger tapping. Additionally, musicians' movements consistent with the instrument of training were less variable and more synchronous than finger tapping. This suggests that movement abilities may be relatively specialized to the motor effector with which musicians are trained. These findings raise questions about factors that might influence a central timing mechanism and the degree to which motor learning may be domain-specific.

Working Memory and Auditory Imagery Predict Synchronization With Expressive Music

Ian Colley[1], Peter Keller[1], Andrea Halpern[2]
[1]The MARCS Institute for Brain, Behaviour, and Development, Western Sydney University, Australia
[2]Department of Psychology, Bucknell University, USA

Sensorimotor synchronization (SMS) is prevalent and readily studied in musical settings, as most people are able to perceive and synchronize with a beat (e.g. by finger tapping). We took an individual differences approach to understanding SMS with real music characterized by expressive timing (i.e. fluctuating beat regularity for artistic interpretation) in a musically untrained sample. Given the dynamic nature of SMS, we hypothesized that individual differences in working memory and auditory imagery—both fluid, cognitive processes—would predict SMS at two levels: 1) asynchrony (a measure of overall synchronization error), and 2) anticipatory timing (i.e. predicting, rather than reacting to fluctuating beat interval sizes). In Experiment 1, non-musician participants completed two working memory tasks, four measures of auditory imagery (vividness of imagery, control of imagery, pitch imagery, and tempo imagery), and an SMS-tapping task. Hierarchical regression models were used to predict SMS performance, with results showing dissociations among imagery types in relation to asynchrony, and evidence of a role for working memory in anticipatory timing. In Experiment 2, a new sample of participants completed an expressive timing perception task to examine the role of imagery in perception without action. Results suggest that imagery vividness and imagery control (the ease with which one changes an image) are important for perceiving and synchronizing with, respectively, irregular but ecologically valid musical time series, whereas working memory is implicated in strategically synchronizing by anticipating events in the series.

Perception of auditory and visual disruptions to the beat and meter in music

Jessica E. Nave-Blodgett[1], Erin E. Hannon[1], and Joel S. Snyder[1]
[1]University of Nevada, Las Vegas. Las Vegas, Nevada, USA

Humans can perceive an isochronous beat in rhythmic auditory patterns such as music, and use this information to make perceptual judgments, tap to the beat, and dance. Music theory suggests there is a hierarchy of beats, with some stronger (measure downbeats) and others weaker (upbeats). Do listeners regularly perceive these so-called metrical hierarchies in music or in simple auditory and visual patterns? Adult musicians and non-musicians listened to excerpts of ballroom dance music paired with either auditory (beeps) or visual (clock-face) metronomes. Participants rated how well the metronome fit the music they heard. The metronomes could be synchronous with the music at the level of the beat but not the measure, the measure (the strongest beat in the repeating groups) but not the beat, both the beat and measure, or neither, yielding four metronome conditions. Sessions were blocked by metronome modality. Overall, participants rated beat synchronous auditory and visual metronomes as fitting the music better than beat-asynchronous metronomes. Musicians showed evidence of hierarchical perception, rating all measure and beat synchronous metronomes as fitting the music better than beat synchronous metronomes in both modalities. Non-musicians showed the same pattern of ratings as musicians, but their ratings did not differ as strongly between fully synchronous and beat (but not measure) synchronous metronomes in either modality. Thus all listeners successfully extracted beat-level information from musical excerpts and matched it to the auditory and visual metronomes. For both groups beat-level synchrony outweighed measure-level synchrony, but musicians were more strongly influenced by measure-level synchrony than were non-musicians. Formal training in music may enhance sensitivity to these additional levels of organization. Without a need to use these hierarchical beat levels in their daily lives, non-musicians may not be as sensitive to higher metrical levels.

Metricality modifies the salience of duration accents but not pitch accents

Jon B. Prince[1], Tim Rice[1]
[1]Psychology and Exercise Science, Murdoch University, Perth, WA, Australia

Background
Prince (2014) showed that tonality enhances the salience of pitch leap accents at the expense of duration accents in a metrical grouping task, even though tonality was irrelevant to the task and provided no grouping cues. Thus the mere presence of organisational structure in pitch (i.e., tonality) enhanced its dimensional salience (Prince, Thompson, & Schmuckler, 2009). However it is not known whether temporal organisational structure can similarly enhance the salience of unrelated temporal information.

Aims
The aim was to investigate if metricality (whether a sequence conforms to a regular metre) affects the salience of duration accents.

Method
Baseline study (N=29): participants rated on a 5-point scale how strongly each atonal 24-note sequence evoked a duple (accent every two notes) or triple (accent every 3 notes) grouping. Pitch and duration accents were tested separately in order to determine values that gave equivalent levels of accent strength. Both metric and ametric (random IOI) sequences were tested. Experimental study (N=32): same procedure, but sequences contained both pitch and duration accents.

Results
Overall, pitch accents were less influential than duration accents. Metricality did not affect the strength of pitch accents but it did increase the strength of duration accents.

Conclusions
The overall advantage of duration accents over pitch accents conflicts with earlier findings using this methodology (Ellis & Jones, 2009; Prince, 2014), perhaps caused by including a metric manipulation. Metricality did not decrement the salience of pitch accents, which does not correspond with previous research showing the converse relation – that tonality decremented the salience of duration accents (Prince, 2014). However, metricality enhanced the salience of duration accents, even though the two were entirely independent manipulations, supporting the dimensional salience theory.

Pupillary and eyeblink responses to auditory stimuli index attention and sensorimotor coupling

Lauren K. Fink[1,2], Brian K. Hurley[1,3], Joy J. Geng[1,2,3] and Petr Janata[1,2,3]
[1] Center for Mind and Brain, University of California Davis, Davis, CA, USA
[2] Neuroscience Graduate Group, University of California Davis, Davis, CA, USA
[3] Department of Psychology, University of California Davis, Davis, CA, USA

Behavioral, electrophysiological, and neuroimaging studies have all confirmed that attentional processing and motor output can become entrained to and enhanced by rhythmic auditory stimuli. However, most of these studies have used simple, isochronous stimuli and have rarely assessed auditory attention via eye-tracking. This study aims to 1) determine the effects of complex rhythmic stimuli on visual motor behavior 2) test the power of Tomic & Janata's (2008) oscillator model in predicting both perceptual thresholds and continuous attentional state (as indexed by pupil dilation). Five different repeating 2.2 s, multi-instrument percussion patterns were used as stimuli. In each stimulus, a deviant could occur at four possible time points: two time points corresponding to moments of high, model-predicted attentional salience and two to low. An infrared eye-tracker (Eyelink 1000, sampling at 500Hz) was used to record eye movements. Participants (n = 18; mean age: 26 +/- 8) were instructed to comfortably maintain their gaze on a fixation cross, attend to the rhythmic patterns, and press the spacebar anytime they heard an increase in volume (200ms intensity deviant). Each of the five patterns looped for approximately 7 minutes; an adaptive thresholding algorithm adjusted the intensity of each subsequent deviant at each probe position. We found that pupil diameter oscillated in a stimulus-specific manner and was a reliable index for distinguishing detected vs. undetected deviants. The likelihood of eyeblink occurrence also systematically mapped onto stimulus structure. As expected, perceptual thresholds for detecting deviants at moments of high-predicted salience were low (better performance), and vice versa. This study highlights the model's ability to predict both behavioral and neurophysiological data, as well as the value in using eye-tracking, a non-invasive technique, in music research.

Rhythmic timing in the visual domain aids visual synchronization of eye movements in adults: Comparing visual alone to audio-visual rhythms

Melissa Brandon[1], Nicole Willett[1], Diana Chase[1], Dillon Lutes[1] and Richard Ravencroft[1]
[1]Psychology, Bridgewater State University, Bridgewater, MA, USA

Past research on adults' inability to synchronize to visual rhythms has been used as evidence for the assumption that audition is for temporal processing and vision is for spatial processing (Repp & Penel, 2002, 2004; Patel et al., 2005). But recent research has shown adults can entrain to environmental visual rhythms (Richardson, et al. 2007). Likewise, adults show improved tapping synchronization to visual rhythms that include motion (Hove, Spivey, & Krumhansl, 2010; Iversen et al., 2009). This recent research suggests when the visual rhythms are ecological the visual system does accurately respond to the temporal information. The aim for this series of studies was to examine whether adults can make use of visual temporal information to synchronize eye-movements to visual rhythms. In one study adults' eyes were tracked watching Sesame Street characters appear in a constant spatial pattern with either rhythmic or jittered timing (between-subjects). An image of a cloud was present in the center of the screen but unobtrusive to the visual rhythm, and in the second half the cloud moved to occlude a location in the rhythmic pattern. First looks to each location in the rhythm were classified as anticipatory or reactive and the response times recorded. The adults used the rhythm to plan anticipatory eye-movements that were more accurate in time compared to the jittered condition, $F(1,30)=7.55$, $p=0.01$. The placement of the cloud did not affect visual synchronization. The anticipatory looks were closer to the onset of the characters in the rhythmic condition compared to the jittered condition. These findings demonstrate that adults do track visual temporal information and use it to plan eye-movements. In a follow-up study, adults' eyes were tracked watching audio-visual rhythms (a tweeting bird) to examine the impact of the additional auditory information on synchronization. Preliminary results show similar synchronized eye-movements for the audio-visual rhythm.

Motor system excitability fluctuations during rhythm and beat perception

Daniel J. Cameron[1], Jana Celina Everling[1], T.C. Chiang[1], Jessica A. Grahn[1]
[1]Western University, Ontario, Canada

Background
Humans synchronize movements with the perceived, regular emphasis (the beat) in musical rhythms. Neural activity during beat perception is dynamic, time-locked, and heavily based in the motor system. Neural oscillations synchronize to regularities, such as the beat, in auditory rhythms (Fujioka, et al., 2012; Nozardan, et al., 2011). Excitability in the motor system fluctuates during beat perception, as indexed by the amplitude of motor-evoked potentials (MEPs) evoked by transcranial magnetic stimulation (TMS). In a previous study, MEP amplitude was greater when listeners heard rhythms with a strong beat vs. a weak beat, but only for MEPs elicited 100 ms before the beat, *not* for MEPs generated at random time points relative to the beat (Cameron, et al., 2012). Thus, motor system excitability may not be uniform across the entire rhythm, but instead may fluctuate, rising at particular times on or before the beat.

Aims
We sought to characterize the dynamics of motor system excitability during beat perception.

Methods
Participants (n=23) listened to strong beat, weak beat, and nonbeat auditory rhythms (35s each). TMS was applied to left primary motor cortex at 100 time points in the beat interval of each rhythm type. MEP amplitudes were recorded with electromyography from the right hand (first dorsal interosseous muscle).

Results
Motor system excitability dynamics differed for strong beat vs. weak beat and nonbeat rhythms: Excitability ncrease linearly over the beat interval (in anticipation of upcoming beat positions) in strong beat rhythms, but not weak or nonbeat rhythms. Additionally, excitability fluctuated at the rate of individual events (the beat rate subdivided by four) to a greater extent in strong beat rhythms than in the other rhythms.

Conclusions
These results suggest that during beat perception, motor system excitability 1) anticipates upcoming beats, and 2) tracks subdivisions of the beat.

Does anticipating a tempo change systematically modulate EEG beta-band power?

Emily Graber[1], Takako Fujioka[1]
[1]Department of Music, Center for Computer Research in Music and Acoustics, Stanford University, USA

Neural oscillatory activities are known to relate to various sensorimotor and cognitive brain functions. Beta-band (13-30Hz) activities are specifically associated with the motor system; they are characterized by a decrease in power prior to and during movement (event-related desynchronization, ERD), and an increase in power after movement (event-related synchronization, ERS). Recent research has shown that listening to isochronous auditory beats without motor engagement also induces beta-band power fluctuations in synchrony with the beat; the rate of ERS is proportional to the beat rate suggesting that ERS may be related to temporal prediction. Top-down processes such as imagining march or waltz meters also induce modulations related to the temporal structure of the particular meter. Here we hypothesize that beta-band modulations may reflect top-down anticipation in other timing-related tasks that commonly occur in music performance. Specifically, we examine whether actively anticipating a tempo change can be differentiated into anticipating an increasing, a decreasing, or a steady tempo. Electroencephalographic (EEG) signals were recorded from musicians who were asked to actively anticipate tempo changes in auditory beat stimuli. Prior to each trial, an informative visual cue appeared on a screen to indicate that the beat rate would increase, decrease, or remain steady. The initial beat onset interval was 600ms for at least 6 beats before a gradual tempo change began. The onset of the tempo change was randomized across trials. Using time-frequency analysis, beta-band power modulations were examined in the EEG data segment just before the tempo changes occurred. Preliminary results replicate earlier findings of beta-band ERD and ERS in synchrony with the beat rate. Further analysis will investigate the mean power, modulation depth, and various electrode sites to differentiate the three conditions.

An EEG Examination of Neural Entrainment and Action Simulation during Rhythm Perception

Jessica Ross[1], John Iversen[2], Scott Makeig[2], and Ramesh Balasubramaniam[1]
[1] Cognitive and Information Sciences, University of California, Merced, Merced, CA, USA
[2] Swartz Center for Computational Neuroscience, Institute for Neural Computation, University of California, San Diego, La Jolla, CA, USA

Listening to musical rhythms activates our motor system, including premotor cortex, supplementary motor area, cerebellum and increases corticospinal excitability as measured by Transcranial Magnetic Stimulation (TMS) and electromyography. According to the Action Simulation for Auditory Prediction (ASAP) hypothesis this motor activation while listening to rhythmic sounds could be internal motor simulation, which in turn helps the auditory system predict upcoming beats in rhythmic sounds. In the present experiments we recorded EEG from participants who were instructed to listen to rhythmic music while not moving their body. Preliminary data show evidence for neural entrainment in beta band EEG and in mu wave suppression. We are presently using Independent Components Analysis (ICA) and dipole source localization to determine the motor sources of rhythm-related modulation of EEG. The results are discussed in the context of the role of motor simulation even in the absence of explicit movement.

The BEAT test: training fundamental temporal perceptual motor skills

Joel Afreth[1], Scott Makeig[1], and John R. Iversen[1]
[1]Swartz Center for Computational Neuroscience, Institute for Neural Computation, University of California, San Diego, La Jolla, CA, USA

A growing body of work suggests connections between rhythmic abilities and more general cognitive performance. An outstanding issue is if training in complex beat perception (involving the perception and maintenance of a periodic pulse in the face of highly syncopated or ambiguous rhythms) is in any way more effective than mere synchronization with a metronome. We have designed an adaptive test and training method for enhancing beat perception and production, with two ultimate goals: 1) examining if this kind of training yields measurable benefits to attention and other cognitive behaviors, and 2) examine if real-time brain computer interface (BCI) feedback might augment training. We will describe the Beat Expertise Adaptive Test (BEAT), which assesses the limits of an individual's ability to maintain a beat to complex rhythms. In the BEAT participants synchronize to the beat of a continuously evolving rhythm starting with a strong, clear beat (simple march-like rhythms easy to synchronize to) and evolving into patterns with a weakly supported beat (syncopated patterns with no dominant beat). Successful synchronization (defined by absolute asynchrony over the preceding 3 beats < 100 ms) causes the rhythms to become progressively harder, while poor synchronization causes the current rhythm to become easier, enabling the adaptive determination of an individual complex beat synchronization threshold. We will present initial BEAT results for nine participants pre- and post- a 30 min interactive beat perception training program that expands the principles of the BEAT into a mutli-level game with visual feedback. Subjects varied in performance both before training and in their relative change after training, with the majority showing significant gains in maximal difficulty achieved after training. We will present preliminary analyses of simultaneously recorded EEG, testing if it is possible to develop a signature for loss of synchronization.

Detrended Cross-correlation Analysis Reveals Long-range Synchronization in Paired Tempo Keeping Task

Masahiro Okano,[*1] Masahiro Shinya,[*2] Kazutoshi Kudo[*3]

[*]*Department of Life Sciences, The University of Tokyo, Japan*

[1]masahiro.okano.sr@gmail.com, [2]shinya.masahiro@gmail.com, [3]kudo@idaten.c.u-tokyo.ac.jp

ABSTRACT

Skilled musicians can synchronize their sound timings each other while performance tempo of music involves random fluctuation. Some recent studies have conducted synchronization task with stimulus sequence having random variability and found evidence of complexity matching and long-range cross-correlations (LRCC) between stimulus sequence and response sequence. While keeping both intended tempo and synchronization with partners is a fundamental skill for musical ensemble, previous studies have not examined whether the complexity matching and/or LRCC is found in paired tempo-keeping synchronization tapping task without external timing cue: this would enhance our knowledge about dynamics of musical synchronization. We conducted a paired tempo keeping synchronization tapping task. Significant complexity matching and cross-correlation about inter-tap intervals were revealed in time scales of 20 seconds or longer whereas they were not observed in the shorter time scales. These results indicate that the interpersonal synchronization would contain multiple coordination processes in different timescales. Complexity matching and DCCA can be interesting tools for investigating musical synchronization.

I. INTRODUCTION

The tempo of live music performance involves random fluctuations; yet, we can synchronize with others in an ensemble. There is much evidence that humans can synchronize their actions with stimulus sequences having variable inter-stimulus intervals (Delignières & Marmelat, 2014; Stephen, Stepp, Dixon, & Turvey, 2008; Stepp & Turvey, 2010). In such cases, complexity matching, that is, scaling exponents of response and stimulus sequences have been found to become approximately the same (Delignières & Marmelat, 2014; Stephen et al., 2008). The complexity reflects the long-range autocorrelation (fractal structure) of the time series and is derived from the slope of power spectrum density (PSD) or detrended fluctuation analysis (DFA) (Delignières & Marmelat, 2014). Also, detrended cross-correlation analysis (DCCA), a recently developed method, enables us to search long-range cross-correlation (LRCC) between time-series (Delignières & Marmelat, 2014; Hennig, 2014; Podobnik & Stanley, 2008).

Previous studies found complexity matching in various task such as synchronization with chaotic metronome (Stephen et al., 2008), pendulum swinging (Fine, Likens, Amazeen, & Amazeen, 2015; Marmelat & Delignières, 2012), synchronization with chaotic arm movement (Washburn, Kallen, Coey, Shockley, & Richardson, 2015), syncopated finger tapping (Coey, Washburn, Hassebrock, & Richardson, 2016). However, no studies examined whether complexity matching is found in tempo-keeping synchronization tapping task without external timing cues. Keeping steady tempo without a metronome and keeping synchronization with the partners are fundamental skills for musical ensemble. Complexity matching is interpreted as reflecting global coordination between participants, not simply local timing adjustment (Stephen et al., 2008). To examine whether complexity matching is found in tempo-keeping synchronization tapping task without external timing cues would enhance our knowledge about musical synchronization.

Also, LRCC has been found by DCCA in various time series such as stock markets, road traffic, climate, or electroencephalography (Delignières & Marmelat, 2014). However, few studies have applied DCCA to human behavioral data today. The only study we found applying DCCA to paired synchronization tapping task in human was Hennig (2014). Hennig (2014) found LRCC in inter-tap interval (ITI) time series in paired synchronization tapping task. In Hennig (2014), participants were not required to keep initial tempo: therefore, whether LRCC appears when participants intend to keep initial tempo is an open question. LRCC is expected to be a potential effective tool to search global coordination structure between time-series. To apply DCCA to tempo-keeping synchronization tapping task without external timing cues would provide hints to understand musical synchronization.

This study examined the correlation structures of the paired tempo keeping task (paired version of synchronization−continuation paradigm) using PSD, DFA, and DCCA. Applying these methods to a task resembling ensemble playing would allow us to investigate the dynamics of music synchronization in long timescales which one could not have measured by only the mean and variability of asynchronies or ITI.

II. METHODS

A. Participants

Twenty-six (13 pairs) healthy adults (mean age = 24.2 years, SD = 2.2 years) participated in the experiment. Two participants reported expertise of musical instruments (both of them were amateur drummers). All participants provided informed consent for their participation in the experimental procedures, which were conducted in accordance with the Declaration of Helsinki and approved by the ethics committee of the Graduate School of Arts and Sciences of the University of Tokyo.

B. Apparatus

Two electric drum kits (WAVEDRUM Mini, KORG, Tokyo, Japan) were used to measure participants' tap timing and produce sound feedback. Participants tapped the WAVEDRUM sensors (a 9.3 cm × 6.5 cm hard plastic board was pasted on top for easy tapping) in front of them, and the corresponding voltage was sampled at 1,000 Hz, using a 16-bit analog-to-digital converter (USB-6218 BNC, National Instruments, Austin, TX, USA). Participants could hear the metronome beats through a speaker in front of them, and their drum sound and that of their partner were played through the speakers. The metronome beats were presented using an application (Metronome Beats 2.3., www.stonekick.com) on the smartphone (ZenPhone5, ASUS) and sampled at 1,000 Hz using the same analog-to-digital converter. Partitions were placed between the participants in each pair and the experimenter, to avoid visual contact and possible distraction resulting from the experimenter's gaze.

C. Paired Tempo Keeping Task

In our experiment of paired tempo keeping task, participants were required to tap in synchrony with each other as well as maintain the reference tempo as accurate as possible, as if they were playing the role of a member of an ensemble. In each trial, participants tapped in synchrony with metronome beats at first. Ten seconds after the participants started tapping, the metronome was turned off, and they continued tapping, keeping as close as possible to the reference tempo. The reference tempos were set at 75 bpm, 120 bpm, and 200 bpm (800 ms, 500 ms, and 300 ms in ITI, respectively). These settings were intended to cover slow, medium, and fast musical tempo. The order of reference tempos was counterbalanced. All the participants performed the task twice.

D. Analyses

1) *Tap Onset Detection.* We preprocessed data to optimize tap onset detection, using the following procedure: We obtained the zero-mean voltage time series by subtracting the mean value from the original time series, and then full-wave rectified that. Data were passed through a 100–300 Hz second-order Butterworth band-pass filter. Filtered data were Hilbert transformed, and the instantaneous amplitude was calculated to obtain an envelope of the waveform. The tap onset times in the preprocessed data were defined as the times at which the value was larger relative to that of the threshold, and the value of the adjacent 50-ms window was smaller relative to that of the threshold.

2) *Detrended Cross-correlation Analysis (DCCA) and Detrended Fluctuation Analysis (DFA).* DCCA works on two time series $x(t)$ and $y(t)$ of equal length N. Fisrt, we get $X(t)$ and $Y(t)$ by calculating the cumulative sum of deviation time series of $x(t)$ and $y(t)$, respectively, for each time point t.

$$X(t) = \sum_{i=1}^{t}[x(i) - \bar{x}] \text{ and } Y(t) = \sum_{i=1}^{t}[y(i) - \bar{y}] \quad (1)$$

These integrated series are divided into k non-overlapping segments of length n. and the series $X(t)$ and $Y(t)$ are then locally detrended, and the covariance of detrended $X(t)$ and $Y(t)$ is calculated as

$$F^2_{\mathrm{DCCA}}(n) = \frac{1}{N}\sum_{t=1}^{N}[X(t) - X_n(t)][Y(t) - Y_n(t)] \quad (2)$$

where $X_n(t)$ and $Y_n(t)$ denote theoretical values obtained from regression of $X(t)$ and $Y(t)$, respectively. This calculation is repeated over all possible interval lengths n. In typical, the longer interval length n, $F^2_{\mathrm{DCCA}}(n)$ more increases. A power law is expected:

$$F^2_{\mathrm{DCCA}}(n) \propto n^{2\lambda} \quad (3)$$

λ is the scaling exponent and expressed as the slope of log-log plot of $F_{\mathrm{DCCA}}(n)$ as a function of n. When we input $x(n)$ to equation (1) instead of $y(n)$, DCCA reduct to DFA:

$$F_{\mathrm{DFA}}(n) = \sqrt{\frac{1}{N}\sum_{t=1}^{N}[X(t) - X_n(t)]^2}. \quad (4)$$

$$F_{\mathrm{DFA}}(n) \propto n^{\alpha} \quad (5)$$

Zebende (2011) noted that λ itself did not reflect the strength of cross-correlation, and introduced DCCA cross-correlation coefficient $\rho_{\mathrm{DCCA}}(n)$, defined as

$$\rho_{\mathrm{DCCA}}(n) = \frac{F^2_{\mathrm{DCCA}}(n)}{F_{\mathrm{DFA}_x}(n)F_{\mathrm{DFA}_y}(n)}, \quad (6)$$

where $F_{\mathrm{DFA}_x}(n)$ and $F_{\mathrm{DFA}_y}(n)$ represent $F_{\mathrm{DFA}}(n)$ for $x(t)$ and $y(t)$, respectively. $\rho_{\mathrm{DCCA}}(n)$ is a set of coefficients which vary between -1 (perfect anti cross-correlation) to 1 (perfect cross-correlation) and represents the evolution of cross-correlation with n. $\rho_{\mathrm{DCCA}}(n) = 0$ means no cross-correlation in the time-scale of interval length n.

In DFA, the value α = 0.5 indicates the absence of autocorrelations, that is, the input signal was white noise. α > 0.5 indicates persistent long-range correlations, that is, large (small) values are more likely to be followed by large (small) values. α < 0.5 indicates anti-persistent autocorrelations, meaning that large values are more apt to be followed by small values and *vice versa*. The values α = 1 and α = 1.5 means the input signal was 1/*f* noise and Brownian noise, respectively (Peng, Havlin, Stanley, & Goldberger, 1995). The relationship between DFA α and slopes derived from PSD (-β) is α = (β+1) / 2 (Delignieres, Ramdani, Lemoine, Torre, Fortes, & Ninot, 2006). It should be noted that α itself does not reflect the entire property of long-range autocorrelation, either: Some task provide $F_{\mathrm{DFA}}(n)$ plot with different slope between different time scales (Peng et al., 1995; Riley, Bonnette, Kuznetsov, Wallot, & Gao, 2012).

We applied DFA and DCCA to ITI time series obtained from paired tempo-keeping synchronization task. We also calculated PSD because fractal analyses are recommended to perform in a combination of multiple methods (Delignieres et al., 2006). All these three methods were performed after removing linear trend because drift of time series unreasonably increases low-frequency power of PSD and correlations in longer time scales in DFA and DCCA.

III. RESULTS and DISCUSSION

One participant was unable to synchronize with their partner; therefore, the data from the pair including this participant were excluded from subsequent analyses. The following results are based on the data from the remaining 12 pairs (24 participants). Figure 1 demonstrates typical examples of $F_{\mathrm{DFA}}(n)$ and PSD plots. As shown in figure 1, the slopes of long time scales (> 20 s) were about the same, whereas those of short (< 20 s) were different. Figure 2 demonstrates correlations of PSD slopes between participants in each pair. In short time scales, the slopes of participants in a pair were not correlated ($r(24)$ = -.11, .36, and -.23 for 75, 120, and 200 bpm, respectively, ps > .05), whereas those were highly correlated in

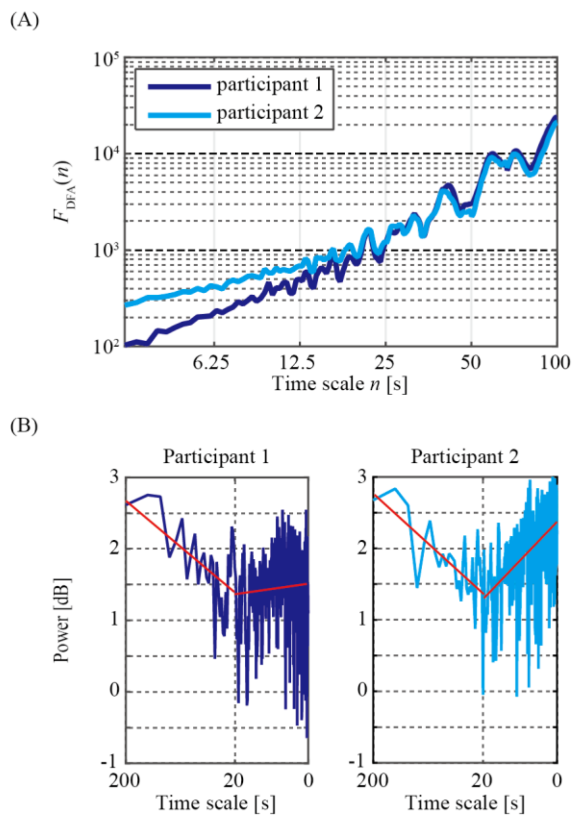


Figure 1. (A) A typical example of $F_{DFA}(n)$ of ITI time-series in a trial. (B) Power spectrum of ITI time-series in the same trial. In both plots, the slopes of longer time scale (> 20 s) were about the same, whereas those of shorter (< 20 s) were different.

long time scales (r (24) = .93, .96, and .75 for 75, 120, and 200 bpm, respectively, ps < .001), suggesting complexity matching in long time scale. Also, differences of the correlation coefficients between long and short time scales were

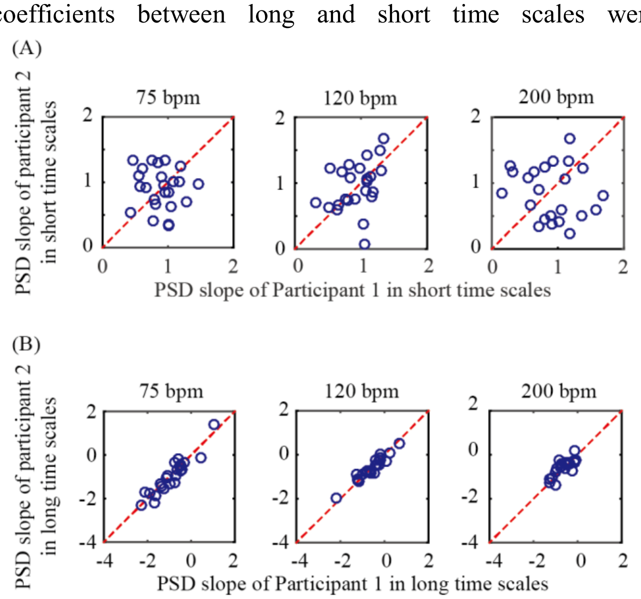


Figure 2. Correlations of PSD slopes between participants in each pair (A) for short time scales (< 20 s) and (B) for long time scales (> 20 s). In short time scales, the slopes of participants in a pair were not correlated, whereas those were highly correlated in long time scales.

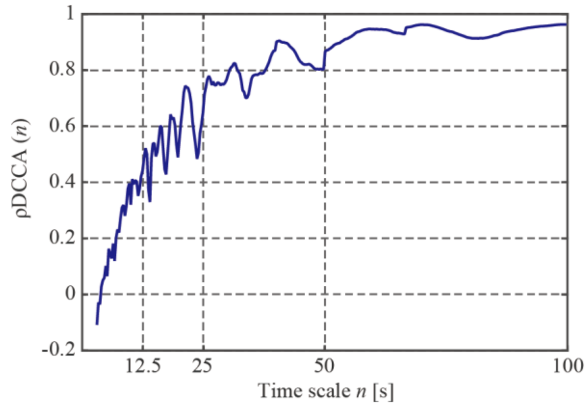


Figure 3. A typical example of ρDCCA (n) in a trial (the same as of figure 1). The greater correlations were obtained for the longer time scales.

significant (z = 5.84, 4.91, and 3.94 for 75, 120, and 200 bpm, respectively, ps < .001). Complexity matching has been discussed as reflecting participant's *strong anticipation* (Stephen et al., 2008; Stepp & Turvey, 2010). *Strong anticipation* is a form of anticipation that is not based on internal models (Stepp & Turvey, 2010). For example, a system made up of two coupled external semiconductor lasers, of which the transmitter laser takes a delayed feedback input from itself whereas the receiver takes an input from the transmitter at negligible delay, demonstrates anticipating synchronization: measured beam intensity of the receiver over time shows similar waveform to that of the transmitter and the receiver wave shows phase lead (Stepp & Turvey, 2010). In an experiment by Washburn et al. (2015), participants (coordinators) could synchronize their arm movements to chaotic arm movements of their partners (producers) even with delayed feedback. Coordinators with delayed feedback should anticipate the producers' movement to synchronize with, whereas the movements of the producers were confirmed to be chaotic one by Lyapunov exponent. This type of anticipation could not be explained by the internal model (anticipation based on internal models is referred to as *weak anticipation* in contrast to *strong anticipation*). Complexity matching in long time scales would suggest participants' ability to sense and adapt global structures of partner's tapping.

Figure 3 demonstrates a typical example of ρDCCA (n), suggesting that the greater correlations were obtained for the longer time scale. While Hennig (2014) demonstrated that global coordination emerges from local coordination, our result suggests that global coordination does not necessarily need local coordination. No study has reported such a type of dynamics as long as we know; further investigation is needed to understand the process that causes global coordination without local one.

In conclusion, PSD, DFA and DCCA enable us to investigate the temporal coordination property in various time scales. These methods would be interesting tools to understand dynamics of musical synchronization.

ACKNOWLEDGMENT

This study is partially supported by the research grant from Tateishi science and technology foundation awarded to MO and Grants-in-Aid for Scientific Research (KAKENHI) No. 24240085 awarded to KK.

REFERENCES

Coey, C. A., Washburn, A., Hassebrock, J., & Richardson, M. J. (2016). Neuroscience Letters Complexity matching effects in bimanual and interpersonal syncopated finger tapping. *Neuroscience Letters*, *616*, 204–210.

Delignières, D., & Marmelat, V. (2014). Strong anticipation and long-range cross-correlation : Application of detrended cross-correlation analysis to human behavioral data. *Physica A*, *394*, 47–60.

Delignieres, D., Ramdani, S., Lemoine, L., Torre, K., Fortes, M., & Ninot, G. (2006). Fractal analyses for 'short'time series: a re-assessment of classical methods. *Journal of Mathematical Psychology*, *50*(6), 525-544.

Fine, J. M., Likens, A. D., Amazeen, E. L., & Amazeen, P. G. (2015). Emergent Complexity Matching in Interpersonal Coordination : Local Dynamics and Global Variability. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(2), 1–15.

Hennig, H. (2014). Synchronization in human musical rhythms and mutually interacting complex systems. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(36), 12974–12979.

Marmelat, V., & Delignières, D. (2012). Strong anticipation: complexity matching in interpersonal coordination. *Experimental Brain Research*, *222*(1-2), 137–48.

Peng, C. K., Havlin, S., Stanley, H. E., & Goldberger, A. L. (1995). Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos (Woodbury, N.Y.)*, *5*(1), 82–87.

Podobnik, B., & Stanley, H. E. (2008). Detrended Cross-Correlation Analysis: A New Method for Analyzing Two Nonstationary Time Series. *Physical Review Letters*, *100*(8), 084102.

Riley, M. A., Bonnette, S., Kuznetsov, N., Wallot, S., & Gao, J. (2012). A tutorial introduction to adaptive fractal analysis. *Frontiers in Psychology*, *3*(371), 1–10.

Stephen, D. G., Stepp, N., Dixon, J. A., & Turvey, M. T. (2008). Strong anticipation: Sensitivity to long-range correlations in synchronization behavior. *Physica A: Statistical Mechanics and Its Applications*, *387*(21), 5271–5278. http://doi.org/10.1016/j.physa.2008.05.015

Stepp, N., & Turvey, M. T. (2010). On Strong Anticipation. *Cognitive Systems Research*, *11*(2), 148–164.

Washburn, A., Kallen, R. W., Coey, C. A., Shockley, K., & Richardson, M. J. (2015). Harmony from chaos? Perceptual-motor delays enhance behavioral anticipation in social interaction. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(4), 1166–1177.

Zebende, G. F. (2011). DCCA cross-correlation coefficient: Quantifying level of cross-correlation. *Physica A: Statistical Mechanics and Its Applications*, *390*(4), 614–618.

Eye movement synchronization to visual rhythms in infants and adults

Melissa Brandon[1], Sarah Hackett[1], Jessica Gagne[1] and Allison Fay[1]
[1]Psychology, Bridgewater State University, Bridgewater, MA, USA

The assumption that audition is for temporal processing and vision is for spatial processing has fueled research on adults' inability to synchronize to visual rhythms (Repp & Penel, 2002, 2004; Patel et al., 2005). But recent research shows adults can entrain to environmental visual rhythms (Richardson et al., 2007) and improve tapping synchronization to moving visual rhythms (Hove, Spivey, & Krumhansl, 2010; Iversen et al., 2009). This recent research suggests the visual system does process temporal information. This series of studies used eye-tracking to explore eye responses to visual rhythms, looking for synchronization. Infants and adults were compared to gain a developmental perspective. Eight-month-olds and adults watched Sesame Street characters move in a constant spatial pattern with either rhythmic or jittered timing (between-subjects). First looks to target locations were classified as anticipatory or reactive and the response times recorded. The infants reacted to the characters rather than anticipated their locations. However, in the first half of the experiment infants' looks were significantly faster in the rhythmic than the jittered conditions, $F(1,30)=7.30$, $p=0.01$. The adults showed no difference between the rhythmic and jittered conditions in this simple experiment (*ps* >0.05), but in a 2nd experiment with a distracter object, the adults' looks were significantly faster in the rhythmic than the jittered conditions, $F(1,30)=7.55$, $p=0.01$. The adults' anticipatory looks were closer to the onset of the characters in the rhythmic condition. These results indicate infants and adults do track visual rhythms and use them to plan their corresponding eye-movements. Follow-up studies using audio-visual rhythms with adults will be discussed for the impact of the addition of auditory information. Finally, future directions for a contingent looking paradigm based on visual synchronization accuracy will be proposed as an applied methodology.

Neural entrainment during beat perception and its relation to psychophysical performance

Molly J. Henry[1] and Jessica A. Grahn[1]
[1]Department of Psychology, Brain and Mind Institute, The University of Western Ontario, London, ON, Canada

The ability to pick up on regularities in environmental stimuli is apparent in infancy and supports language learning, movement coordination, and parsing auditory scenes into "objects". Here, we were interested in the seemingly unique sensitivity humans show to temporal regularities in rhythm: they spontaneously feel a "beat" in rhythmic sequences. In particular, we examined how synchronization of neural oscillations with auditory rhythms might give rise to beat perception, and in turn how entrained neural oscillations might affect psychophysical performance. In the current electroencephalography (EEG) study, participants detected near-threshold targets (changes in spectral bandwidth) embedded in long (~19-s) simple or complex auditory rhythms. Simple rhythms were composed of intervals related by integer ratios (1:2:3:4), and had a regular grouping that resulted in standard and target events always being present at "on-beat" locations given a duple meter. Simple rhythms thus induced a relatively strong sense of a beat. Complex rhythms were also composed of intervals related by integer ratios, but were grouped irregularly and thus did not induce a strong beat percept. We compared spectral power at beat-related frequencies (1.25, 2.5, and 5 Hz, where 5-Hz was the base inter-tone interval) for simple and complex rhythms. We observed significantly stronger spectral power at 1.25 Hz for simple compared to complex rhythms, indicating stronger subharmonic entrainment for rhythms that gave rise to a strong sense of beat. We did not observe power differences between rhythm types at 2.5 or 5 Hz. Ongoing analyses will reveal the relations between entrained neural phase/power and perception, and how these relations depend on the presence of a perceived beat. The results will for the first time link electrophysiological correlates of beat perception with psychophysical consequences of sensing a beat.

Deconstructing nPVI

Nathaniel Condit-Schultz[1]

[1]*School of Music, the Ohio State University, USA*

ABSTRACT

The normalized pairwise variability index (nPVI) is a mathematical measure originally used to characterize the rhythm of languages. Patel and Daniele (2003) introduced the nPVI to music research, and it has since been fruitfully used in a number of papers. This paper presents a methodological criticism of the nPVI as applied to music. It is argued that that simpler mathematical characterizations, which are more musically intuitive, can capture the same useful information as nPVI. Specifically, counting the proportion of adjacent IOIs that are identical accounts for as much as 98% of variation in nPVI scores in symbolic musical corpora.

I. INTRODUCTION

The *normalized pairwise variability index* (*nPVI*) is a mathematical measure of durational contrast between adjacent rhythmic events, originally used to characterize the rhythm of languages (Grabe and Low, 2002). Patel and Daniele (2003) introduced the nPVI to music research, finding that a difference in the average nPVI between spoken English and French was parallelled in instrumental themes by English and French composers (Patel et al., 2006)—a result which was quickly replicated by Huron and Ollen (2003) using a much larger dataset. Huron and Ollen (2003) found that themes by composers from a variety of additional European nationalities also evinced differences in average nPVI. McGowan and Levitt discovered similar correlations between musical and linguistic nPVI across three English dialects (McGowan and Levitt, 2011).

Though introduced as a tool for studying language/music parallelism, nPVI has since seen been used to study musical rhythm in general. For instance, Patel and Daniele (2013; 2015) discovered historical trends in the nPVI scores of Western music, both across and within composers. Several scholars have directly investigated the usefulness of nPVI as applied to music: Toussaint (2012) empirically tested the correlation between nPVI and a number of alternative mathematical characterizations of musical rhythm (Toussaint, 2012). London and Jones (2011) explored several refinements to the nPVI measure, including incorporating grouping boundaries and studying higher level rhythms.

The nPVI is an abstract, dimensionless quantity, falling in the range $(0, 200]$. The advantage of this dimensionless quality—which attracted Patel and Daniele (2003, p. B37)—is the affordance of comparison across different rhythmic domains (i.e., music and language). However, this abstractness is also a liability, as it makes it difficult to interpret or imagine the actual musical relevance of variation in nPVI. What exactly these nPVI values reflect?; what musical difference is captured by, for instance, the nPVIs of 46.9 and 40.9 observed in English and French themes by Patel and Daniele (2003, p. B41)? Is this a large difference, or a small one? All that can be said for certain is that higher nPVI values reflect greater durational contrast between adjacent events.

Due to its abstract nature the nPVI formula can be applied to any form of ordered durational data, including musical or linguistic rhythms. However, there are fundamental differences between the rhythmic organization of language and music, differences with serious implications for the interpretation of musical nPVIs. Specifically, musical durations within a given piece are drawn from a small set of durations which form simple whole-number ratios with each other, such as the 2:1 ratio between ♪ and 𝅘𝅥𝅯. This is especially the case when nPVIs are calculated from "quantized" durational data (for instance, traditional music notation), as has been the case in all the studies mentioned above.

This paper will present a simple critique of the application of nPVI to music. The argument is not that the nPVI fails to capture useful information. Rather, the thesis is simply that the organization of musical rhythm into simple whole-number ratios affords simpler mathematical characterizations which capture the same useful information as the nPVI, but which are more musically intuitive.

II. DATASET

This paper draws upon four musical corpora as datasets to explore and illustrate the nature of nPVI: (1) The European and (2) Chinese components of the Essen database of folk music; (3) The first violin part from a convenient sample of

ISBN 1-876346-65-5 © ICMPC14

58 Haydn string quartets[1]; (4) The author's corpus of popular rap transcriptions, the Musical Corpus of Flow (MCFlow).[2] Figure 1 shows the distribution of nPVI scores across various subsections of the four corpora. The variation in nPVI evident in Figure 1 is broadly consistent with the results reported in other research. For instance, the scope of variation in nPVI between European regions is comparable to the scope of regional variation observed by Huron and Ollen (2003). However, between the nine regional categories shared between the Essen dataset and Huron and Ollen's dataset, the nPVI values are show no significant correlation. In fact, the previously observed difference between French and English instrumental themes is dramatically reversed in the Essen dataset, with nPVIs of 52.9 and 45.5 respectively.

III. DISCUSSION

A. Calculating nPVI

To calculate an nPVI the following equation is applied to each adjacent pair in an ordered series of inter-onset-intervals (IOIs). Each pairwise calculation results in a value in the range $[0, 200)$, which I will refer to individually as *normalized Pairwise Values* ($nPVs$); The nPVI is simply the mean of these nPVs.

$$nPV = 200 * \left| \frac{\text{antecedent IOI - consequent IOI}}{\text{antecedent IOI + consequent IOI}} \right| \quad (1)$$

B. Interpreting nPVs

There is a simple one-to-one relationship between nPVs and pairwise IOI ratios, as illustrated in Figure 2.[3] Such IOI-ratios are generally more intuitive than abstract nPVI values, especially to musicians. For example, the relationship between ♩ and ♪ is more easily understood as the ratio 2:1 than as the nPV 66.67.

The absolute value signs in Equation 1 make the order of the IOIs irrelevant, which in ratio terms results in reciprocal equivalence: i.e. 1:2 = 2:1; In more musical terms, ♪→♩ is considered the same as ♩→♪. The rational nature of nPVs also means that the absolute size of IOIs is irrelevant; for example, ♩→♪ and ♪→𝅘𝅥𝅯 are equivalent—this is the "normalization" in nPVI. The principle reason for this normalization is to remove the influence of absolute speed, and changes of speed, in speech. However,

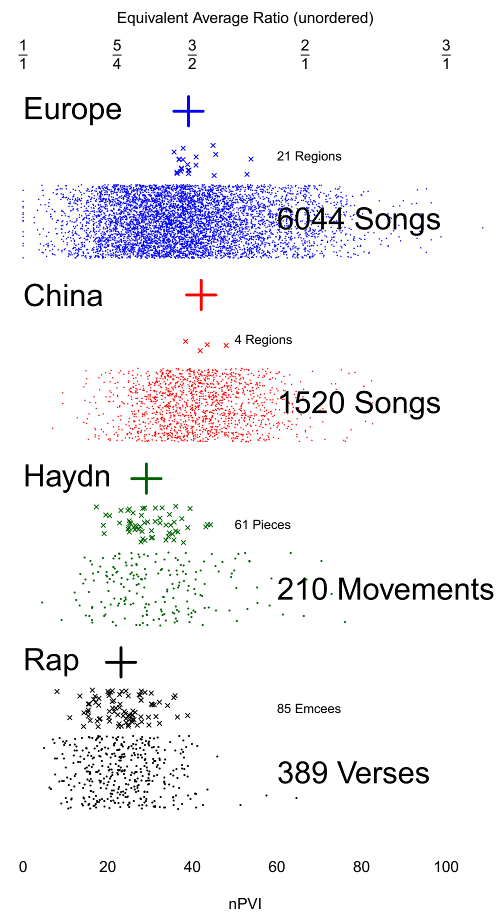


Figure 1: Distribution of nPVI scores across four corpora. For each corpora, the corpus' global nPVI is marked with a large cross-hair, subregions are indicated by small xs, and individual songs/movements are marked with small dots. nPVI values are randomly "jittered" across the y-axis so that individual points are visible.

musical notation does not contain any absolute durational information, so the effect of this normalization in the analysis of music notation is very different than in the analysis of speech.

C. Distribution of nPVs

Figure 3 shows a histogram of nPVs within each corpus. Notice that, though nPVs are in principle continuous, when applied to symbolic music notation the practical result is a completely categorical distribution, with nearly all pairwise relationships forming the ratios 1:1, 2:1, 3:1, or 4:1—corresponding to nPVs of 0, 66.67, 100, or 120 respectively. If durations were extracted from human music performances these precise categories would blur. Still, given the nature of musical rhythm, we would still expect the pairwise ratios found in live musical performance to be heavily spiked near these ideal whole-number ratios. Figure 4 show similar histograms for five individual

[1] The Essen and Haydn datasets were accessed in kern format through the Kern Scores website (kern.humdrum.org).

[2] MCFlow is freely available at rapscience.net.

[3] The relationship between IOI ratios and the nPV is captured by the function $f(x) = 200 * \left|\frac{x-1}{x+1}\right|$.

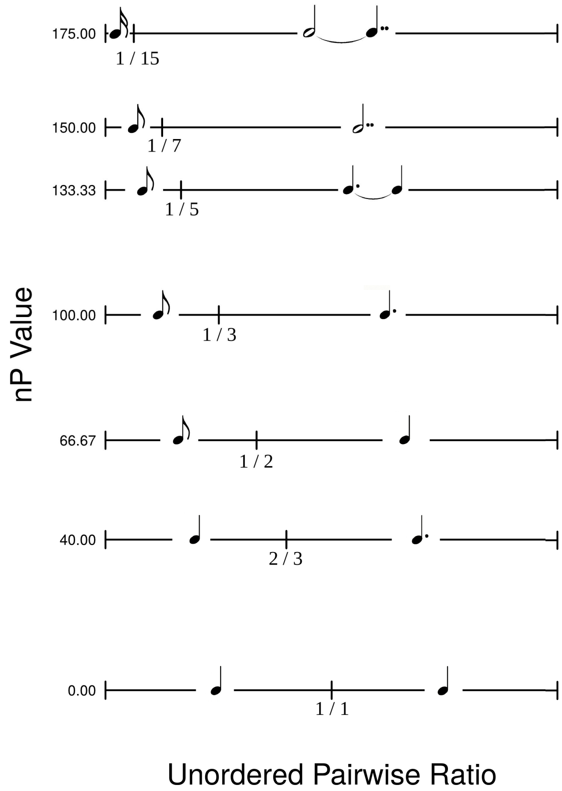


Figure 2: Relationship between IOI ratios and nPVI. The ratios should be interpreted as unordered, meaning that 1:2 receives the same nPV as it's reciprocal 2:1.

songs drawn from the European corpus. The five songs were chosen to represent a range of nPVI scores within the European corpus. Specifically, the songs' nPVIs represent the 10%, 30%, 50%, 70%, and 90% quantiles of the European corpus. In other words, the first song's nPVI is greater than only 10% of the European songs', while the last song's nPVI is greater than 90% of the European songs'.

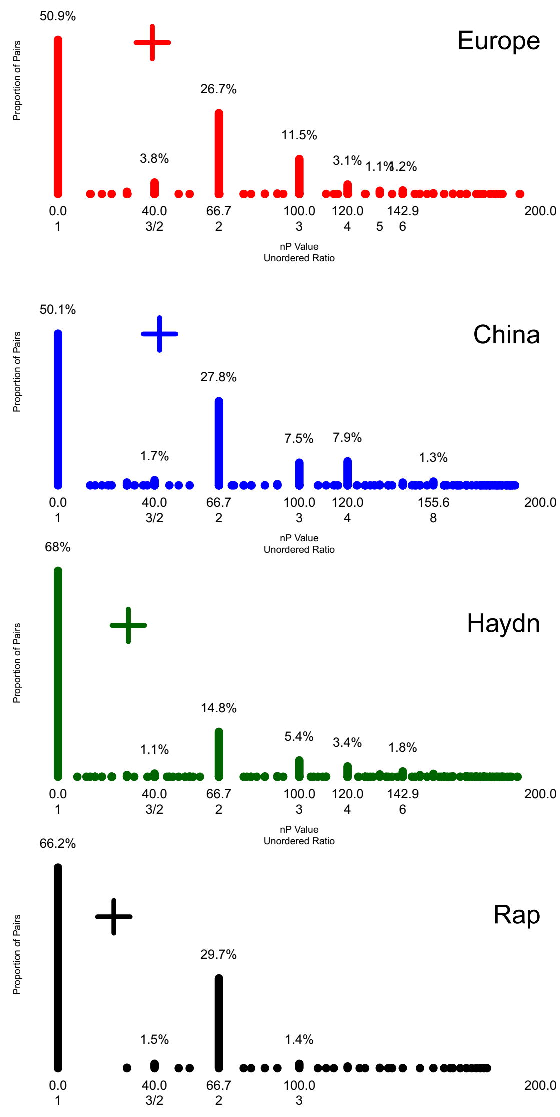


Figure 3: Histograms of nPV in each of four corpora. Large cross marks the average of the values, which is the nPVI.

D. Characterizing nPVs

An nPVI is simply an average of nPVs; the nPVI of each nPV-distribution is marked with a cross symbol in Figures 3 and 4. However, a simple mean is not necessarily the most useful descriptor of such categorical distributions. How might we better characterize such distributions? Jian (2004) proposed using the median nPV rather than the mean. However, the median of all the distributions shown in Figures 3 and 4 are all 0, except the last (> 90%) song, with a median value of 66.67. Similarly, the modes of all nine distributions are 0. Thus, neither the mode nor median seem to be as sensitive as the mean in detecting rhythmic differences.

One striking feature of Figures 3 and 4 is the concentration of 1:1 pairs. This reflects the highly regular, periodic nature of musical rhythm. In fact, it appears that much of the information in these distributions is simply captured by the proportion of 1:1 ratios ($nPV = 0$). To test this informal observation, a simple linear regression model was created to predict nPVI scores for each song in each of the four corpora using the "proportion of 1:1 pairs" as a predictor. The resulting adjusted R^2 for the four corpora are Europe = .89, China = .84, Haydn = .88, and Rap = .98. Thus, 84–98% of variance in nPVI is accounted for simply by the proportion of repeated IOIs. This suggests that the answer to the question "what aspect of musical rhythm does nPVI capture?" is largely "how often IOIs repeat." If the proportion of 2:1 ratios ($nPV = 66.667$) are added to this model, the adjusted R^2 values are

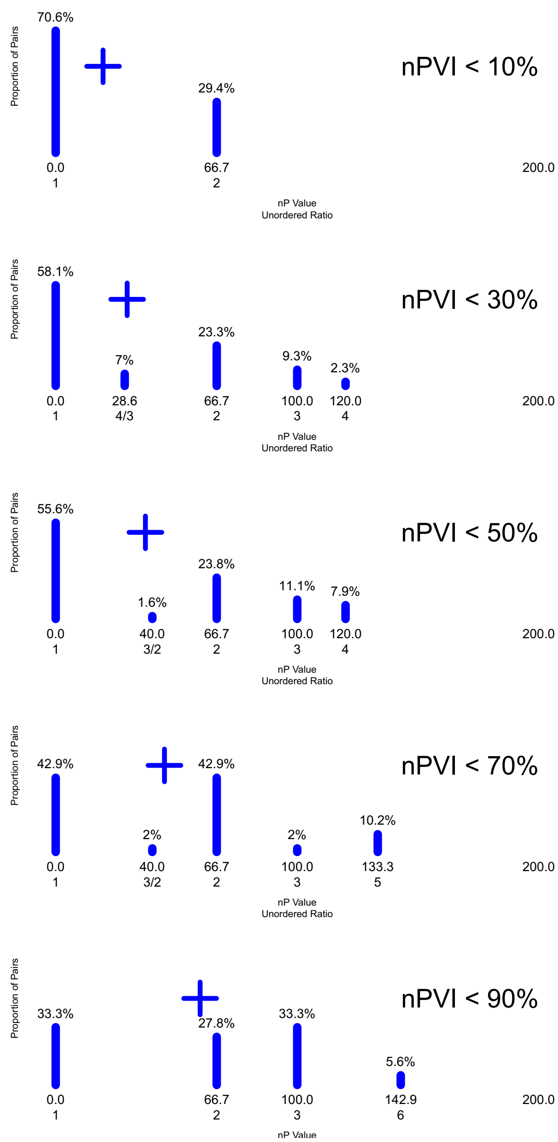


Figure 4: Histograms of nPVs in five songs drawn from the European corpus. The five songs were selected based on the position of their nPVI score within the distribution of nPVI scores in the European corpus. This the top histogram describes a song with a relatively now nPVI score which is greater than only 10% of the European songs. The middle graph plots a song whose nPVI score is greater than 50% of nPVI scores in the European corpus (the median). A cross marks the average of the values (the nPVI).

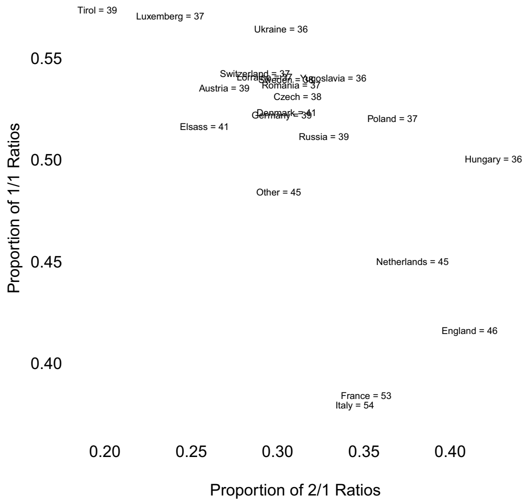
raised to Europe = .98, China = .96, Haydn = .98, and Rap = .98.

It is often convenient to reduce complex information like the distributions shown in Figures 3 and 4 to a single number, as nPVI does. However, it may ultimately be more fruitful to retain more information. For instance, by comparing the nPV distributions directly, we can more thoroughly explore the differences in pairwise durational relationships between French and English. The proportion of 1:1 pairs in French and English songs are 41.5% and 38.3% respectively—a fairly minor difference. However, the biggest difference between French and English IOI pairs, is that French songs in the corpus contain approximately 63% more 3:1 ratios than English songs. Interestingly, this ratio forms rhythms such as ♩. ♪, which are associated with the opening section of *French* overtures. Only by studying the complete distribution of pairwise ratios, can more precise observations such as this be made.

As a compromise between a single index value and the complete nPV distribution, we can report a 2–4 dimensional "pairwise durational profile." For instance we could present the proportion of 1:1, 2:1, 3:1, and 4:1 ratios in the data, which seems to account for the vast majority of pairs. Even a two-dimensional profile may be significantly more informative than a one-dimensional index. Figure 5 shows two two-dimensional plots of the twenty one European regions from the Essen corpus, showing the relative proportion of 1:1, 2:1, and 3:1 IOI pairs.

IV. CONCLUSION

The principle conclusion of this paper is that the abstract, dimensionless nPVI can effectively be replaced with the more intuitive "proportion of 1:1 pairs" in most research. Beyond this, more complex multidimensional consideration of pairwise IOI ratios may ultimately be more fruitful than any one-dimensional measure. Ultimately, an important goal for future research will be determining which mathematical characterizations of durational contrast best represent the aesthetic intuitions of music listeners. In the process, other assumptions of the nPVI need to be tested. For instance, the unordered nature of nPV values (2:1 = 1:2), and whether or not "normalization" of musical durations is really appropriate. In addition, non-pairwise analyses—perhaps n-grams of 3–4 IOIs—could be considered as well.



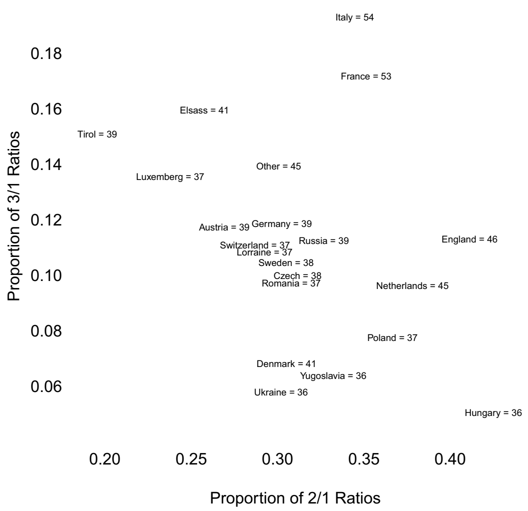


Figure 5: Top figure: the 21 European regions in the Essen corpus plotted by their proportion of 1:1 and 2:1 IOI pairs. Bottom figure: the 21 European regions in the Essen corpus plotted by their proportion of 3:1 and 2:1 IOI pairs. In both figures, regions' nPVI values are printed next to their names.

REFERENCES

Daniele, J. R. and Patel, A. D. (2015). Stability and change in rhythmic patterning across a composers lifetime: A study of four famous composers using the npvi equation. *Music Perception: An Interdisciplinary Journal*, 33(2):255–265.

Grabe, E. and Low, E. L. (2002). *Durational variability in speech and the rhythm class hypothesis*, pages 515–546. Berlin: Mouton de Gruyter.

Huron, D. and Ollen, J. E. (2003). Agogic Contrast in French and English Themes: Further Support for Patel and Daniele (2003). *Music Perception*, 21(2):267–271.

Jian, H.-L. (2004). An improved pair-wise variability index for comparing the timing characteristics of speech. *Proceedings of the 8th International Conference on Spoken Language Processing*, pages 4–8.

London, J. and Jones, K. (2011). Rhythmic Refinements to the nPVI Measure: A Reanalysis of Patel & Daniele (2003a). *Music Perception*, 29(1):115–120.

McGowan, R. W. and Levitt, A. G. (2011). A Comparison of Rhythm in English Dialects and Music. *Music Perception*, 28(3):307–314.

Patel, A. D. and Daniele, J. (2003). An empirical comparison of rhythm in language and music. *Cognition*, 87:B35–B45.

Patel, A. D. and Danielle, J. R. (2013). An empirical study of historical patterns in musical rhythm: Analysis of german & italian classical music using the npvi equation. *Music Perception*, 31(1):10–18. Document feature - Graphs; ; Musical Scores; Tables; Last updated - 2014-08-04.

Patel, A. D., Iversen, J. R., and Rosenberg, J. C. (2006). Comparing the rhythm and melody of speech and music: The case of british english and french. *Journal of the Acoustical Society of America*, 119(5):3034–3037.

Toussaint, G. T. (2012). The pairwise variability index as a tool in musical rhythm analysis. In *Proceedings of the 12th International Conference on Music Perception and Cognition*, pages 1001–1008.

Quantifying synchronization in musical duo improvisation: Application of n:m phase locking analysis to motion capture data

Shinya Fujii[1], Masaya Hirashima[2], Hama Watanabe[1], Yoshihiko Nakamura[3], Gentaro Taga[1]
[1]Graduate School of Education, The University of Tokyo, Tokyo, Japan
[2]Center for Information and Neural Networks (CiNet), National Institute of Information and Communications Technology and Osaka University, Osaka, Japan
[3]Graduate School of Information Science and Technology, The University of Tokyo, Tokyo, Japan

The musical beat can serve as a catalyst to elicit spontaneous rhythmic movements, such as the bobbing of our heads. One-to-one phase locking analysis is commonly used to quantify synchronization of the movement-to-the-beat (e.g., Patel et al., 2009; Phillips-Silver et al., 2011; Fujii and Schlaug 2013), yet more complex phase locking patterns can occur in music. The aim of this study was to quantify complex movement synchronization in musical duo improvisation by applying n:m phase locking analysis (Tass et al., 1998). Sixteen professional drummers participated in the study and each of them improvised with a professional bassist. The head movements of the drummer and the bassist were recorded by a motion capture system with a sampling frequency of 200 Hz. The audio tracks were also recorded at 48000 Hz. Sound features (root mean square, spectral centroid, spectral entropy, spectral flux, brightness, pulse clarity, and event density measures) were extracted from the audio tracks with the MIR toolbox (Lartillot and Toiviainen, 2007). Beat times and the inter-beat intervals were determined with a beat tracking algorithm (Ellis, 2007). The motion capture data was segmented with the beat times and bandpass filtered to extract the movement frequencies corresponding to the half, fourth, and eighth note beat frequencies. We then calculated the degree of n:m phase locking of head motion between pairs of players. We found that there were multiple patterns of transition in phase locking mode during the improvisation. For example, one drummer-bassist pair showed 1:1 phase locking mode at the beginning, then changed the pattern into 1:2 phase locking mode in the middle, and finally showed 1:1 phase locking mode again at the end. The pattern of transition in the phase locking mode was matched with the pattern of changes in the sound features. The results suggest that the n:m phase locking analysis is useful as it can reveal the dynamics of complex musical synchronization.

Spared motor synchronization to the beat of music in the presence of poor beat perception

Valentin Bégel[1,5], Charles-Etienne Benoit[1] , Angel Correa[6,7], Diana Cutanda[10], Sonja A. Kotz[8,9], and Simone Dalla Bella[1,2,3,4]
[1]EuroMov, Montpellier University, Montpellier, France
[2]Institut Universitaire de France (IUF), Paris, France
[3]International Laboratory for Brain, Music, and Sound Research (BRAMS), Montreal, Canada
[4]Dept. of Cognitive Psychology, WSFiZ, Warsaw, Poland
[5]NaturalPad, SAS, Montpellier, France
[6]Centro de Investigacion Mente, Cerebro y Comportamiento, University of Granada, Spain
[7]Departamento de Psicologia Experimental, University of Granada[8]Maastricht University, Faculty of Psychology & Neuroscience, Dept. of Neuropsychology & Psychopharmacology, Maastricht, The Netherlands
[9]Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
[10]Centro de Investigacion Mente, Cerebro y Comportamiento, University of Granada, Spain

Moving to the beat of music is a widespread reaction in humans. Even though the majority can synchronize to the beat, some individuals encounter major difficulties in performing this task (i.e., poor synchronizers). Poor synchronization was initially ascribed to poor beat perception (i.e., beat deafness). However, it has been shown that poor synchronization to a beat can co-exist with unimpaired beat perception. In the current study, we present two cases of beat-deaf individuals (L.A. and L.C), who exhibit an opposite dissociation: they can tap to the beat of music while their perception is impaired. In a first experiment, participants' explicit timing skills were assessed with the Battery for The Assessment of Auditory Sensorimotor and Timing Abilities (BAASTA), which includes both perceptual and sensorimotor tasks. Their performance was compared to that of a control group. L.A and L.C performed poorly on rhythm perception tasks, such as detecting durational shifts in a regular sequence. Yet, they could tap to the beat of the same stimuli (i.e., their performance was both accurate and consistent). In a second experiment, we tested whether the aforementioned perceptual deficit in L.A. and L.C. extend to an implicit timing task. In this task the regular temporal pattern of isochronous auditory stimuli aids in performing a non-temporal tasks (i.e., responding as fast as possible to a different target pitch after a sequence of standard tones). L.A. and L.C's performance in the implicit timing task was comparable to that of controls. Altogether, the fact that synchronization to a beat can occur even in the presence of poor perception suggests that timing perception and action may dissociate at an explicit level. This finding is reminiscent of dissociations observed in pitch processing (i.e., in poor-pitch singers). Spared implicit timing mechanisms are likely to afford unimpaired synchronization to the beat in some beat-deaf participants.

Rhythmic Consonance and Dissonance: Perceptual Ratings of Rhythmic Analogs of Musical Pitch Intervals and Chords

Alek S. Razdan,*[1] Aniruddh D. Patel#[2]

* *Depts. of Math and Physics (B.S. 2015), Tufts University, USA*

Dept. of Psychology, Tufts University, USA

[1]alekrazdan@hotmail.com, [2]a.patel@tufts.edu

ABSTRACT

The current study quantifies human preferences for 'rhythmic intervals' (polyrhythms) formed by expressing the frequency ratios of common Western musical pitch intervals and triads as rhythms. We hypothesized that pleasantness ratings of these rhythmic stimuli would resemble those of their pitch counterparts. Consistent with this hypothesis, we find that pleasantness ratings of rhythmic intervals and triads correlate significantly with pleasantness ratings found in earlier research on corresponding pitch intervals and triads. Pleasantness ratings for our polyrhythms were inversely related to ratings of their rhythmic complexity, and positively related to ratings of how strongly they induced a sense of an underlying beat ('beat induction') and how much they made one want to move ('groove'). Future work will aim to determine if neurally-inspired models of rhythm perception based on coupled nonlinear oscillators can be used to explain our findings.

I. INTRODUCTION

In research on consonance and dissonance in music, a long line of influential studies have assessed preferences for different pitch combinations among Western listeners. For example, McDermott et al. (2010) recently measured ratings of pleasantness for all dyadic pitch intervals and several triads common in Western music. However, significantly less work has explored 'rhythmic consonance and dissonance', i.e., the perceived pleasantness of patterns created by combining different rhythmic streams.

We sought to explore human preferences for 'rhythmic intervals' formed by expressing the frequency ratios of common pitch intervals as rhythms. For example, a 'rhythmic perfect fifth' can be created from two superimposed rhythmic metronomes of rates 120 and 80 BPM, in the tempo ratio 3:2. We presented participants with rhythmic equivalents of all the common dyadic pitch intervals and pitch triads found in Western music and asked them to rate each for pleasantness. Inspired by previous work investigating cognitive similarities between pitch and rhythm processing (Pressing, 1983), we hypothesized that pleasantness ratings of rhythmic intervals would resemble those of their pitch counterparts. If true, this would suggest that preferences for rhythmic and pitch intervals may rely on similar cognitive or neural processes. For example, certain timing ratios in sound patterns may have similar effects on the neurophysiological coding of auditory patterns at distinct timescales (slow for rhythm, fast for pitch).

In addition, we sought to identify to what extent pleasantness ratings of our stimuli were related to other perceptual ratings of the same stimuli (such as perceived complexity or groove), and to certain objective measures of stimulus structure, such as mean or standard deviation of inter-onset intervals between events in a rhythmic pattern.

II. METHODS

A. Participants

29 participants (15 female) with self-reported normal hearing completed the study (mean age = 19.9, sd = 1.9). All had at least 5 years of musical training or experience (singing or playing a musical instrument) within the last 10 years. All but three were enrolled as undergraduate or graduate students at Tufts University. All gave informed consent, completed a musical experience questionnaire, and were compensated $10.

B. Stimuli

We created rhythmic equivalents of 12 common musical pitch intervals (e.g., perfect fifth, major second, etc.) and 4 chords (major, minor, augmented, diminished) using pairs or triplets of click trains, respectively, using MaxMSP. We did this by using the frequency ratios of the intervals or chords as they would be realized in Just Intonation to create metronomic pulse trains with corresponding tempi ratios (see Table 1, and compare to the Table in Figure 1B of McDeremott et al., 2010). For example, a 'rhythmic perfect fifth' was made by combining two metronomic pulse trains whose tempi were in the ratio of 3:2, and a 'rhythmic major triad' was made by combining three metronomic pulse trains whose tempi were in the ratio of 6:5:4.

Table 1. Rhythmic Stimuli. Note that in equal-tempered music the ratios of the stated intervals are approximate. In our study, the ratios are exact, as they would be in Just Intonation.

Symbol	Interval	Tempo ratio	Cycle duration (seconds)
O	Octave	2:1	1
m2	Minor second	16:15	8
M2	Major second	9:8	4.5
m3	Minor third	6:5	3
M3	Major third	5:4	2.5
P4	Perfect fourth	4:3	2
TT	Tritone	45:32	22.5
P5	Perfect fifth	3:2	1.5
m6	Minor sixth	8:5	4
M6	Major sixth	5:3	2.5
m7	Minor seventh	16:9	8
M7	Major seventh	15:8	7.5
Maj	Major triad	6:5:4	3
Min	Minor Triad	15:12:10	7.5
Aug	Augmented Triad	25:20:16	12.5
Dim	Diminished Triad	64:54:45	32

For each rhythmic interval or chord (polyrhythm) the fastest click train (corresponding to the largest number in the tempo ratio) was set to 120 BPM, which is near the human spontaneous tapping tempo (McAuley et al., 2006). Thus for example the rhythmic perfect fifth had one click train at 120 BPM and the other at 80 BPM, forming a 3:2 ratio in tempo, while the rhythmic major chord had one click train at 120 BPM, one at 100 BPM, and one at 80 BPM, forming a 6:5:4 ratio.

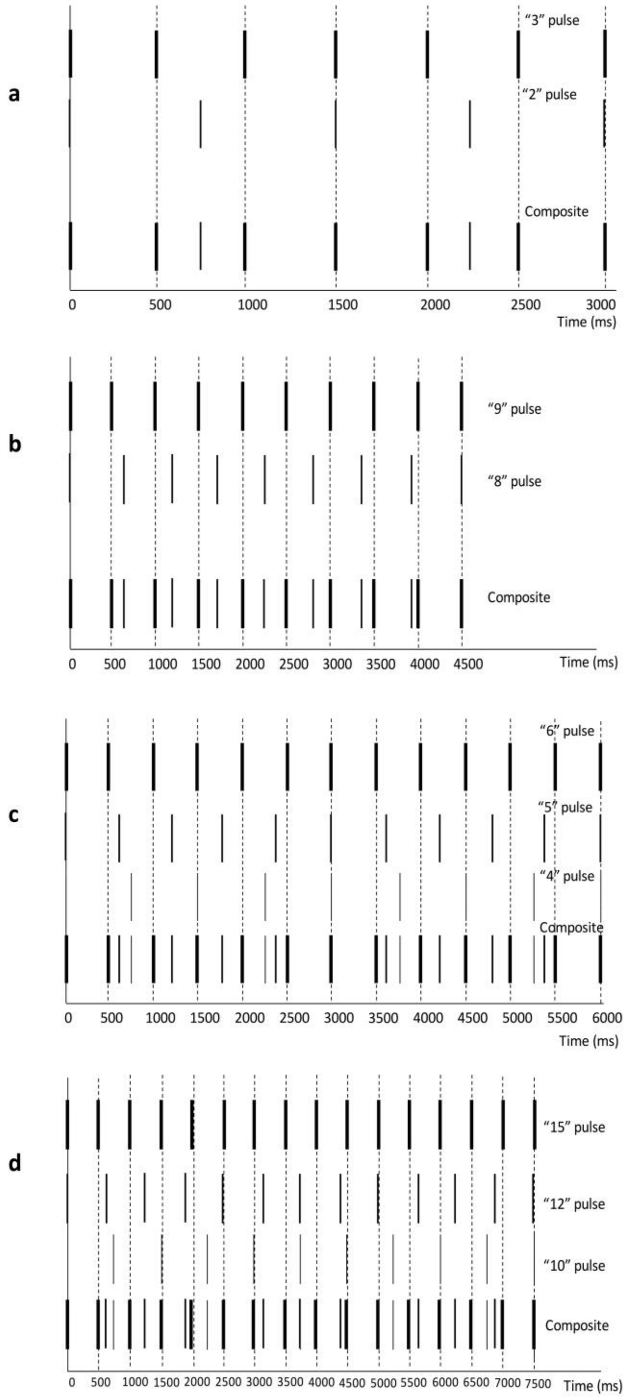


Figure 1. Schematics of four stimuli used in this study. In each panel, the individual metronomic stimuli are shown in the top rows, and the resulting polyrhythm, which is created by adding the streams together, is shown in the bottom row (this was the rhythm heard by participants). a) Rhythmic perfect fifth (3:2 ratio); b) major second (9:8 ratio); c) major triad (6:5:4 ratio); d) minor triad (15:12:10 ratio). In a) and c) two rhythmic cycles are shown; in b) and d) one cycle is shown.

Cycle duration for a given polyrhythm can be computed by multiplying the largest number in the tempo ratio by the period of the fastest click train (always 500 ms), or, alternatively, by multiplying the smallest number in the tempo ratio by the period of the slowest click train. Thus for example, the cycle duration of the rhythmic major sixth was 5 clicks * 0.5s = 2.5 s. Cycle durations for all rhythms are shown in Table 1.

For dyadic polyrhythms, the faster and slower click trains were made using square wave pulses of 4 and 2ms duration, respectively (thus the 3:2 polyrhythm had a 120 BPM train comprised of 4 ms duration square waves with onsets every 500 ms, and an 80 BPM click train comprised of 2 ms duration square waves with onsets every 750 ms). For triadic polyrhythms (e.g., 6:5:4), the fastest click train had 7 ms pulses, the middle-tempo click train had 5 ms pulses, and the slowest click train had 2 ms pulses. A schematic of four of our stimuli is shown in Figure 1.

To create the polyrhythms, click trains were added together, aligned on the first pulse. These polyrhythms were presented diotically over headphones (Behringer HPM1000) in a quiet room, and participants were allowed to adjust the volume to a comfortable level.

C. Task

Since we had 12 dyadic and 4 triadic polyrhythms (cf. Table 1), each participant heard a total of 16 polyrhythms, presented in random order. Rhythms always started on the first beat of the cycle. On each trial participants had to listen for at least 15 seconds, after which they could stop the stimulus at any time to make a rating.

In separate blocks, participants rated the stimuli for pleasantness, complexity, groove, and beat induction (always in that order). Pleasantness was rated on a scale of -3 to 3 (corresponding to the scale used in McDermott et al.'s 2010 study of pitch intervals, with negative indicating unpleasant and positive indicating pleasant). The other ratings were made on a scale of 1 to 7, where 7 was the highest score.

Groove was defined as 'the extent to which the pattern makes you want to move (e.g., tap your foot, bob your head).' Beat induction was defined as the extent to which the pattern projected a clear sense of an underlying beat. Participants took a brief break between each block, and took less than 30 minute to complete all of the ratings.

After the ratings, participants did tests of auditory working memory (forward, backward, and sequencing versions of digit span) using the standardized test from the WAIS IV (Wechsler Adult Intelligence Scale, 4th Edition).

III. RESULTS

Participants found the ratings tasks easy to do. In general, they stopped the stimulus soon after the minimum listening duration of 15 seconds (mean listening time = 18.7 s, sd = 2.3 s). That is, if a pattern took longer than 15s to complete a cycle (cf. Table 1), they did not wait for a cycle to complete before making their rating. This is not surprising since they were not informed that different stimuli had different cycle durations, and for rhythms with long cycles, they may have been unaware that any cyclical structure existed.

Mean pleasantness ratings for our rhythmic stimuli are shown in Figure 2a (see Table 1 for x-axis labels). For comparison, pleasantness ratings of the corresponding pitch intervals are superimposed in Figure 2b (data from McDermott et al., 2010).

As can be seen in Figure 2b, the pattern of pleasantness ratings for our rhythms is similar the ratings of the pitch intervals and chords on which they were modeled (see Appendices A & B for numerical data). The correlation between the ratings in our study and those in McDermott et al.'s study is high ($r = 0.73$, $p = 0.003$). Note that this correlation does not include our pleasantness ratings for the octave or diminished triad, as McDermott et al. did not test these stimuli in their study.

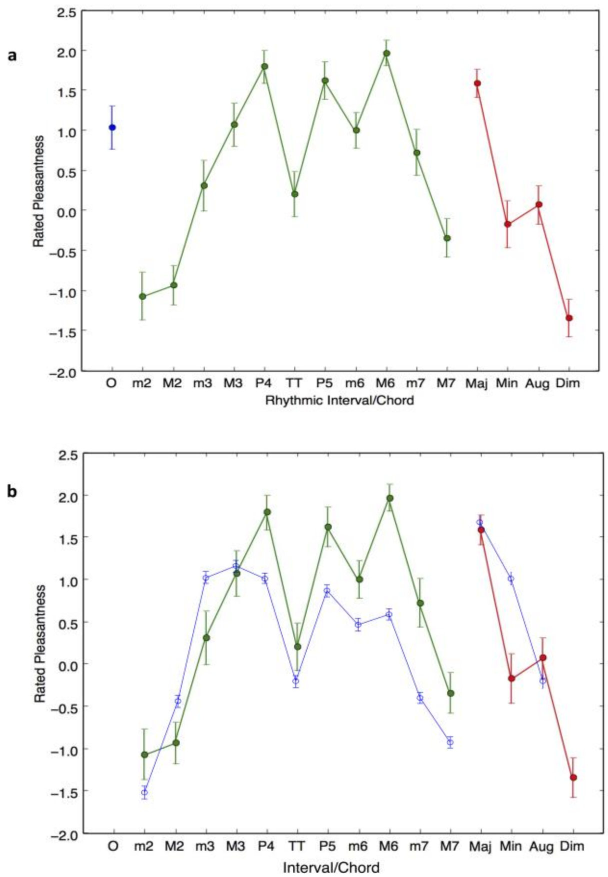


Figure 2. a) Average pleasantness ratings for rhythmic analogs of pitch intervals and chords; b) the same ratings overlaid on the pleasantness ratings of corresponding pitch intervals and chords (blue lines, data from McDermott et al. 2010, synthetic tone condition). McDermott's data did not include ratings for the octave or the diminished triad. In all cases, error bars represent standard errors.

Figure 3 shows means ratings of our rhythmic stimuli for complexity (inverted), groove, and beat induction in a format similar to Figure 2 (see Appendix A for data values). Inverted complexity is simply the complexity data reflected about the median rating scale value of 4 (e.g., a complexity rating of 2 becomes an inverted complexity rating of 6, i.e., higher numbers mean lower complexity: this is done simply to facilitate visual comparison with the pleasantness data). As can be seen by comparing Figure 2a with Figure 3, there is a strong resemblance between the four perceptual ratings (pleasantness, complexity, groove, and beat induction). The pairwise correlations between the mean ratings in each condition are shown in Table 2. (Note that these correlations used complexity, not inverted complexity, for calculation.)

We also sought to determine if the mean pleasantness ratings of each condition correlated with objective aspects of stimulus structure. Specifically, we examined four aspects of each rhythmic pattern: the average and standard deviation of inter-onset-interval durations measured over 1 complete rhythmic cycle; average number of events per second in each cycle, and cycle duration. These correlations were computed across all 16 stimuli, and for the 12 dyadic stimuli only. The results are shown in Table 3.

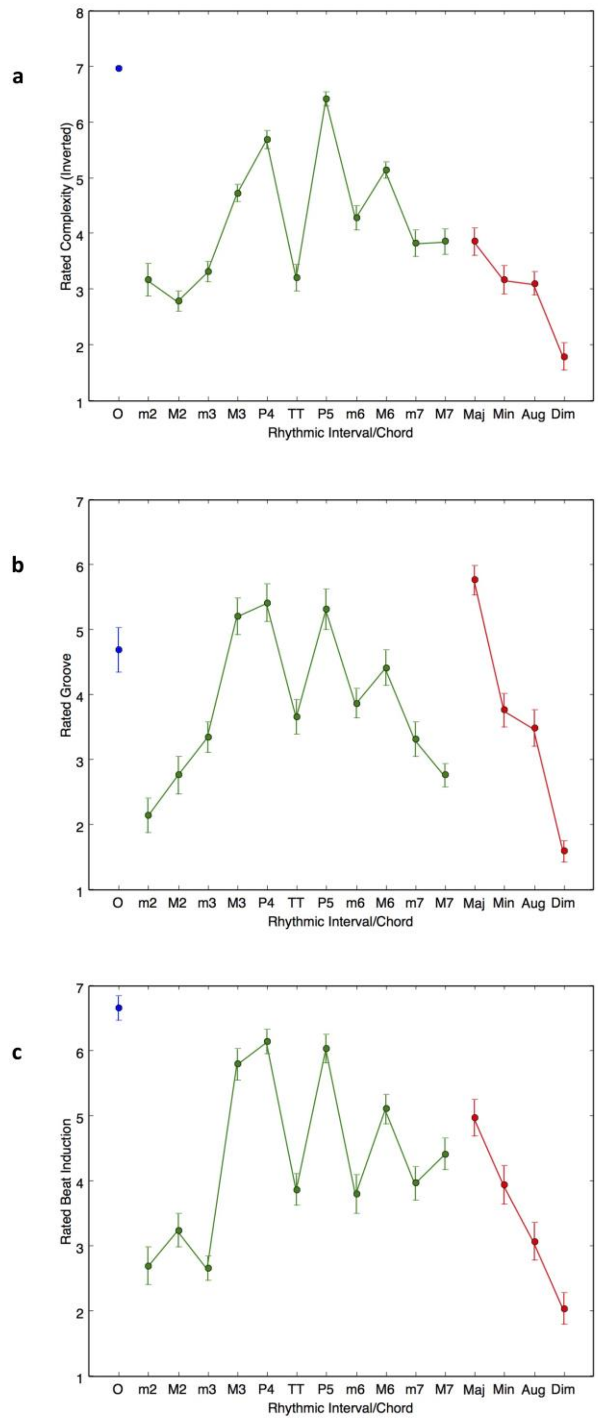


Figure 3. Mean ratings for a) complexity (inverted); b) groove; and c) beat induction for rhythmic analogs of pitch intervals and chords. Errors bars represent standard errors.

Table 2. Correlations between the mean perceptual ratings of pleasantness (P), complexity (C), groove (G), and beat induction (B) for the rhythmic intervals and chords. The table shows r-values for each correlation. For all r-values, $p < 0.001$.

	P	C	G	B
P	--	**-0.78**	**0.91**	**0.79**
C	--	--	**-0.75**	**-0.92**
G	--	--	--	**0.85**

All of the patterns we used involved rhythmic cycles of extended duration, ranging from 1 s for the rhythmic octave to 32 s for the diminished triad (see Table 1). Thus the ability to detect patterns may relate to how much of each pattern could be held in working memory. Being able to hold more of a pattern in mind may help detect structure, and thus lower perceived complexity. To determine if complexity ratings were related to individual differences in working memory capacity, we examined correlations between the average complexity ratings of each participant (across all stimuli) and their measures of working memory capacity (forward, backward, and sequencing digit span). However, none of the results were significant (all $p > 0.1$).

Table 3. Correlations between the mean perceptual ratings of pleasantness and objective measures of stimulus structure. (The octave is excluded from the standard deviation correlations in row 2, as it has a standard deviation of 0 s, an outlier value.)

Correlation of mean pleasantness rating with a rhythm's:	r (all stimuli)	p (all stimuli)	r (dyads only)	p (dyads only)
Average inter-onset interval	**0.526**	**0.003**	0.516	0.086
Standard deviation of inter-onset intervals	-0.175	0.364	-0.463	0.152
Cycle duration	**-0.611**	**$p < 0.001$**	-0.406	0.190
Average number of events per second	**-0.592**	**$p < 0.001$**	**-0.642**	**0.024**

IV. Discussion

Our main finding is that when 'rhythmic intervals and chords' are made by translating frequency ratios of common pitch intervals and chords into metronomic pulse trains with corresponding tempo ratios, the pleasantness ratings of the resulting polyrhythms are remarkably similar to ratings previously reported for the corresponding pitch combinations (Figure 2).

Our ratings of pleasantness are highly correlated with ratings of beat induction and groove, and inversely correlated with ratings of complexity. Thus any of the latter three factors (or some combination of them) could be driving our pleasantness ratings (Table 2). In addition, pleasantness ratings were correlated (or inversely correlated) with some objective measures of rhythmic patterns, such as average inter-onset interval, cycle duration, or average number of events per second (Table 3).

After observing the significant similarity between our pleasantness ratings for polyrhythms and ratings of their pitch counterparts, it is not unreasonable to wonder if a common mechanism may be responsible for driving both of these preference regimes, spanning the temporal spectrum from rhythm to pitch perception. A possible bridge between the timescales of rhythm and pitch may be provided by a framework based on nonlinear oscillatory models of neural activity. Lots and Stone (2008), for example, present a simple model of two neurons coupled in a nonlinear fashion being driven by an external pitch interval stimulus (such as an octave, comprised of two frequencies in the ratio 2:1). They find that the stability of the oscillatory states induced by each of the 12 common pitch intervals in Western music serves as an accurate predictor of its preference ranking.

By stability the authors mean the extent to which, should the intrinsic frequency of one oscillator be perturbed by a small amount, the synchronized relative oscillations of the two coupled oscillators will remain constant. For example, imagine a pair of nonlinearly coupled neural oscillators being driven by an external pitch interval such as the octave (frequency ratio 2:1) or the tritone (frequency ratio 45:32). When the driving interval is the octave, if this driving stimulus is perturbed slightly (say to a stimulus comprised of two pitch streams in the frequency ratio 2.1:1), then the coupled neural oscillators will continue to oscillate in the frequency ratio 2:1. On the other hand, if the oscillators are driven by the tritone, changing the external driving frequency slightly (say, to 46:32) may result in a change of the two oscillators' coupled vibrations from their original frequency ratio of 45:32. Thus, we may speak of the 2:1 oscillatory state as being more stable than the 45:32 state. Intriguingly, Large et al. (2016) show that that if one employs such a nonlinear model and searches for the most stable frequency ratios between 1:1 and 2:1, the 12 common Western intervals found in Just Intonation are effectively reproduced, perhaps providing a neurodynamic explanation for why certain pitch intervals appear repeatedly across many cultures and time periods.

Turning to rhythm, Large & Snyder (2009) and Large et al. (2015) posit that human entrainment to a beat and the tendency to subdivide and accent rhythms may be consequences of nonlinear neurodynamic oscillatory effects. Using a canonical model of nonlinear coupled oscillation, they predict that the most common patterns of metrical accent correspond to stable resonant states of neural oscillation, while the internal sense of a beat that listeners feel could be a direct consequence of spontaneous oscillatory effects inherent in such nonlinear models. Large & Snyder (2009) note that the timescales of rhythm perception and the oscillatory timescales of individual neurons differ significantly from each other. However, they suggest that their canonical model could be applied not just to pairs of individual neurons, but to entire brain regions (specifically, regions associated with both auditory and motor feedback and processing), where activity levels would be on the same timescale as rhythmic perception.

Thus, even though the timescales of pitch and rhythm are quite different, it is possible that the same canonical nonlinear model could be used to explain the perceived pleasantness of related stimuli in both temporal regimes, although the specific regions or neural elements oscillating in each case could be distinct. To help further test such oscillator models, future studies should explore the grey area between pitch and rhythm perception by creating patterns at a range of different tempi. For example, the perfect fifth (ratio 3:2) could be presented using metronomic click trains at 120:80 BPM; 240:160 BPM, 480:320 BPM, etc., up to the realm of pitch at, say, 7680:5120 BPM (=128 Hz:85.3 Hz). One could then study whether pleasantness ratings are consistent between these timescales. If so, this may lend support to such nonlinear oscillator models of rhythm and pitch perception.

Future work should also examine pleasantness ratings in participants with pitch preferences different from those of the Western listeners. For example, some Bulgarian singing traditions prize and hold beautiful certain intervals considered to be highly dissonant by Western listeners, such as the tritone or minor second (Rice, 1980). Asking such subjects to rate the pleasantness of rhythmic and pitch intervals and chords would be a fascinating exercise. If the ratings of pleasantness for pitch and rhythmic intervals and chords were found to be correlated within such a culture, but differed from the ratings given by Western listeners, this would raise questions about how neural oscillator models such as those of Large (2010) can be tuned by cultural experience.

ACKNOWLEDGMENTS

We would like to thank Kameron Clayton for his assistance with presentation of the stimuli and data collection in Matlab, Jennifer Zuk for her help with the training involved in digit span administration, Nori Jacoby for suggestions regarding analysis of objective features of rhythmic patterns, and Josh McDermott for providing the data used in Figure 2b and reproduced in Appendix B. We also thank Michael Hove for his assistance with the creation of the stimuli in MaxMSP.

REFERENCES

Large, E. W. (2010). Neurodynamics of Music. In: M.R. Jones et al. (Eds.), *Music Perception* (pp. 201-231). New York: Springer.

Large, E. W. & Kim, J.C., Flaig, N.K., Bharucha, J.J., & Krumhansl, C.L. (2016). A neurodynamic account of musical tonality. *Music Perception, 33*, 319-331.

Large, E. W. & Snyder, J.S. (2009). Pulse and meter as neural resonance. *Annals of the New York Academy of Sciences, 1169*, 46-57.

Large, E.W., Herrera, J.A., & Velasco, M.J. (2015). Neural networks for beat perception in musical rhythm. *Frontiers in Neuroscience*, *9:*159. doi: 10.3389/fnsys.2015.00159

Lots, I.S., & Stone, L. (2008). Perception of musical consonance and dissonance: an outcome of neural synchronization. *Journal of the Royal Society Interface*, *5,* 1429-1434.

McAuley, J.D., Jones, M. R., Holub, S., Johnston, H. M., & Miller, N.S. (2006). The time of our lives: life span development of timing and event tracking. *Journal of Experimental Psychology: General, 135*, 348-367.

McDermott, J. H., Lehr, A. J., & Oxenham, A. J. (2010). Individual differences reveal the basis of consonance. *Current Biology, 20*, 1035-1041.

Pressing, J. (1983). Cognitive isomorphisms between pitch and rhythm in world musics: West Africa, the Balkans, and Western Tonality. *Studies, in World Music, 17,* 38-61.

Rice, T. (1980). Aspects of Bulgarian musical thought. *Yearbook of the International Fold Music Council, 12*, 43-66.

APPENDIX A

Rating data (mean and standard errors) for pleasantness (P), complexity (C), groove (G), and beat induction (B) for the different rhythmic stimuli in this study. See Table 1 for definition of symbols in column 1. Data are from 29 participants.

Symbol	P	P SE	C	C SE	G	G SE	B	B SE
O	1.034	0.274	1.034	0.034	4.690	0.344	6.655	0.188
m2	-1.069	0.302	4.828	0.290	2.138	0.271	2.690	0.290
M2	-0.931	0.248	5.207	0.182	2.759	0.288	3.241	0.251
m3	0.310	0.314	4.690	0.180	3.345	0.239	2.655	0.188
M3	1.069	0.267	3.276	0.156	5.207	0.282	5.793	0.240
P4	1.793	0.207	2.310	0.165	5.414	0.283	6.138	0.190
TT	0.207	0.282	4.793	0.240	3.655	0.259	3.862	0.242
P5	1.621	0.235	1.586	0.127	5.310	0.310	6.034	0.219
m6	1.000	0.222	3.724	0.221	3.862	0.226	3.793	0.299
M6	1.966	0.161	2.862	0.147	4.414	0.279	5.103	0.229
m7	0.724	0.289	4.172	0.238	3.310	0.268	3.966	0.260
M7	-0.345	0.239	4.138	0.226	2.759	0.177	4.414	0.241
Maj	1.586	0.176	4.138	0.247	5.759	0.231	4.966	0.278
Min	-0.172	0.294	4.828	0.248	3.759	0.251	3.931	0.297
Aug	0.069	0.238	4.897	0.213	3.483	0.283	3.069	0.293
Dim	-1.345	0.234	6.207	0.240	1.586	0.161	2.034	0.240

APPENDIX B

Rating data (mean and standard error) for pleasantness of pitch dyads and triads from McDermott et al. (2010). Data are from the synthetic complex tone condition. Note that McDermott et al. did not collect pleasantness ratings for the octave or diminished triad. Data are from 143 participants.

(NB:These tones were made from a fundamental frequency and 9 upper harmonics in sine phase, with harmonic amplitudes attenuated at 14 dB/octave to mimic naturally occurring musical note spectra. Tone complexes were given temporal envelopes that were the product of a half-Hanning window (10 ms) and a decaying exponential [decay constant of 2.5/sec] that was truncated at 2 s.)

Symbol	Mean pleasantness rating	Standard Error
O	--	--
m2	-1.516	0.074
M2	-0.434	0.069
m3	1.031	0.072
M3	1.166	0.069
P4	1.022	0.065
TT	-0.201	0.069
P5	0.872	0.068
m6	0.471	0.075
M6	0.596	0.068
m7	-0.391	0.067
M7	-0.922	0.070
Maj	1.684	0.075
Min	1.017	0.077
Aug	-0.201	0.079
Dim	--	--

The Effect of Short-term Training on Synchronization to Complex-meter Music.

Carson G. Miller Rigoli,*1 Sarah Creel*2

*Cognitive Science, University of California, San Diego, USA
[1]carsongmr@ucsd.edu, [2]screel@ucsd.edu

I. BACKGROUND

This project examines effects of short-term on complex-meter processing in Western listeners. Previous research shows that North American listeners have difficulty entraining to complex-meter music, producing timing errors which deviate toward simple-meter patterns. While uncommon in the West, complex meter is prevalent in the music of the Balkans and South Asia. This raises the issue of whether the processing difficulties posed by complex meter arise from underlying mechanisms of meter processing or are an effect of Westerners' lack of experience with complex meter.

II. AIMS

Here we examine tap performance to complex 7/8 meters before and after a short training period and discuss the implications of training effects for models of meter processing. We additionally examine the role of experience with Western musical performance on performance with 7/8 meters.

III. METHODS

Participants were asked to tap to drum patterns that corresponded to two 7/8 rhythms common in Balkan music. Tapping accuracy was recorded using an Arduino-based timing device. Participants were trained over the course of a half-hour to tap to one of the two drum patterns. Both drum patterns consisted of three beats with non-isochronous intervals, arranged in interval duration ratios of 3:2:2 or 2:2:3. Each trial included both a synchronization phase in which a drum pattern is presented with a matching melody and a continuation phase in which participants continue tapping the drum pattern without any auditory reinforcement. A questionnaire is also administered to all participants to determine their level of music performance experience as well as their experience with specific musical genres that may indicate familiarity with complex meter. Variability and absolute accuracy of timing were analyzed pre-training and post-training for both patterns in order to determine the amount of distortion toward simple-meter patterns that occurred.

IV. RESULTS

Pre-training results indicated that our participants performed on-par with those reported in a previous study of complex meter tap distortion by Western listeners in both synchronization and continuation conditions. Comparing pre- and post-training performance, we see decreased variability as measured by standard error of tap times in both conditions. There is a significant reduction of distortion with training in the synchronization condition. However, no reduction is seen in the continuation condition and there is no effect of the specific drum rhythm trained on. The musical experience questionnaire was used to classify each participant as a "musician" or "non-musician". It was found that musicians presented significantly more distortion than non-musicians in the synchronization phase prior to training, but made significant improvements and overcame the non-musicians post-training. There was no significant difference by musical experience in learning or absolute performance in continuation.

V. CONCLUSIONS

The observed pattern of results does not well disambiguate between a view of complex-meter distortion as a result of meter-processing which makes use of statistical information or which uses general mechanisms of synchronization. The observed significant improvements made in synchronization but not in the continuation phase as well as the significant difference in synchronization performance by musical experience are however consistent with a model of beat processing in which there is a trade-off in importance of auditory feature extraction and either statistical knowledge of common rhythmic patterns or general beat-tracking processes. In this interpretation of the results, participants begin by relying heavily on endogenously-generated predictions for tap times. As training progresses, they increase the weight of features extracted from the current auditory signal. While these results do not progress the view that predictions for tap times are generated making use of statistical information, they do not contradict that view. In order to better determine underlying mechanisms, future experimentation will involve both Western musicians and non-musicians as well as musicians and non-musicians familiar with complex-meter through immersion in Balkan and North Indian musical cultures.

Do Chord Changes Make Us Hear Downbeats?

Christopher Wm. White[1], Christopher Pandiscio[2]
[1]Department of Music & Dance, The University of Massachusetts, Amherst, Amherst, MA, USA
[2]Department of Electrical & Computer Engineering, The University of Massachusetts, Amherst, Amherst, MA, USA

Musicians have long assumed that harmonic change influences how listeners hear downbeats and meter– our theories and pedagogy have taught that chord changes tend to align with downbeats and metric emphasis (e.g., Lerdahl & Jackendoff 1983, Temperley & Sleator 1999). However, we entertain an alternate explanation. Since humans also perceive strong beats at moments where more *note onsets* (or, pitch initiations) occur (e.g. Boone 2000, Ellis & Jones 2009), it is possible that chord changes are incidental to our perceptions of downbeats– in order to change a chord, more notes must re-strike than when a harmony remains unchanged. It is therefore possible that correlations between harmonic changes and downbeats are due only to patterns of note onsets rather than to the changes in harmonies themselves.

Our experiments tease apart the parameters of harmonic change and note onsets. We present participants with patterns that repeat the same chord twice before changing to a new harmony, but with the first chord of the pair having fewer note onsets than the second. Participants are then asked to tap the stronger pulse. Since the pattern alternates harmonic change with greater numbers of note onsets, if participants tap on the pair's first chord, they are more influenced by harmonic change; if on the second, they are more influenced by note onsets. Our results show that subjects tap on harmonic changes even when the second chord contains 2 or 3 times more note onsets than the first chord; however, subjects tap on the repeated chord when its note onsets are 4-times that of the first chord. Further experiments show related effects of volume, key, and consonance. We end by speculating that our results could arise from an exposure effect: using a corpus of tonal music, we find that offbeats have on average 4-times fewer note onsets than do downbeats, suggesting that this parameter effects listeners only when it conforms to the properties of actual music.

The influence of tempo on subjective metricization

Ève Poudrier[1]
[1]School of Music, University of British Columbia, Vancouver, British Columbia, Canada

Subjective metricization, the parsing of sequences of undifferentiated auditory signals in a wide range of tempi in groups of 2, 3, or 4 qualitatively distinct elements, is one of the earliest phenomenon to have been investigated experimentally (e.g., Bolton 1894, Harrell 1937, Woodrow 1909). Although faster tempi tend to give rise to somewhat larger groups, the effect of rate of presentation appears to be limited. The present study seeked to replicate and extend these findings through online experimentation using a spontaneous grouping paradigm. The specific design was motivated by the limitations of earlier studies, including response bias created by demand characteristics. Three variables were tested: grouping (none, duple, and triple), rate (inter-onset intervals = 200, 550, and 950 ms), and accent (intensity, duration, and pitch). Response type was defined as 1 to 12 elements per group, with 1 representing no grouping. Preliminary findings suggest that in undifferentiated sequences, spontaneous grouping is asymmetrical and tempo-dependent, with 1 and 4 being the most frequent responses within each rate. When every second or third element is differentiated (duple or triple grouping), all three variables have a significant main effect on group length, with intensity accents, triple grouping, and slower rate resulting in longer groups. Most interestingly, although duple and triple grouping generally yielded responses consistent with purely duple (2, 4, and 8) and triple (3 and 9) metric types, compound meters (6 and 12) were most frequently perceived in stimuli with triple grouping, a finding that suggests that lower-level triple grouping gives rise to binary preference at higher levels (3×2 and 3×4), an interpretation that is consistent with the relatively low frequency of response type 9. Overall, rate of presentation also had an intriguing effect on the distribution of responses, with fastest rate of presentation resulting in more even distribution.

"Not On The Same Page" — A Cross Cultural Study on Pulse and Complex-Meter Perception

Hsiang-Ning Dora Kung[#1], Udo Will[#2]

[#] *School of Music, Ohio State University, USA*

[1]kung.34@osu.edu, [2]Will.51@osu.edu

ABSTRACT

Most western meter theories are based on a hierarchical model, with pulse as the basic unit. However, these theories have recently been challenged by the notion that pulse and meter might be processed distinctively. It has also been noticed that the perception of meter may be culturally subjective. The hierarchical model widely accepted in the Western musical academia is probably not appropriate to describe meter perception in other cultures. Through a behavioral experiment, this study explores the relationship between pulse, periodicity and metrical structure from a cross-cultural perspective, and asks whether this relationship is influenced by one's previous metrical experience in a particular musical culture. It also inquires whether there are behavioral indicators for pulse and meter perception to be processed distinctively. If so, how are the two related to each other? From the experiment, no evidence for hierarchical processing of musical meter was found. Instead, the preliminary analyses showed a distinction between pulse, periodicity and meter. The cultural influences, although not addressed much in the current report, can not be ignored in the discussions of rhythmic perception.

I. INTRODUCTION

Today's research of music perception and cognition is principally Western based due to its pioneers' insufficient experience with and understanding of non-Western musical traditions. It has resulted in a common practice in this field that cultural consideration is either excluded or simply ignored in the experimental design. In most of the cases, both the researchers and the recruited participants are restrictively of Western culture or Western musical background that accesses music differently from many other cultures in the world. (Will & Turow, 2011) Given that metrical understanding is a kind of cognitive construction, as opposed to direct percept (Clayton, 2000), metrical interpretation is assumed to be shaped by one's previous musical experience and to be socio-culturally specific. Drawing generalizing conclusions from Western meter theories can therefore be problematic.

Most Western metrical theories are based on hierarchical models that use pulse as its referential unit (Cooper & Meyer, 1960; Lerdah & Jackendoff, 1983). However, in the last couple of years the relationship between pulse and meter has started being challenged by both music theorists and neuroscientists. Fitch (2013) proposed that pulse perception entails the *extraction* of constant pulsations from a series of sound events, while meter perception involves *grouping* the events into hierarchical structures. In an EEG study Kuck et al. (2003) found that although having largely overlapping networks, rhythm and meter are processed distinctively. In another EEG experiment, Geiser et al. (2009) also found that meter processing requires particular attention and is more influenced by musical training than rhythm processing is, suggesting that the former may be more culturally dependent whereas the latter is more likely perceived universally.

For the same reason, the hierarchical models are found inadequate in the ethnomusicological discussions. One of their fundamental issues is that the Western (musical) culture is predominantly nurtured in a literary tradition where abstract cognitive construction is made possible. Without being translated into the visual-spatial domain, however, the acoustic information would be processed sequentially under the constrains of the psychological present that is usually about two to three seconds long, only. Therefore, the aforementioned hierarchical models seem hardly applicable in the cultures where music is primarily, if not solely, communicated through oral transmission. (Cross, 2003; Will & Turow, 2011). On Central African rhythm, for example, Arom (1991) suggested that it is ill-suited to describe it as "metrical" because these repetitive patterns are interlocked around one single pulse level without any implication of strong-weak hierarchy. On the other hand, Clayton (2000) argued that even in cultures having solid metrical theories, like in India, people may listen to the rhythm by grouping similar components according to Gestalt principles, not in metrical ways. He went on to point out that most of the Indian rhythmic cycles (*tal*) are longer than the common duration of psychological present. Although it is not well understood at the moment how *tal*s with a span longer than the psychological present are processed, there is no doubt that one would perceive and understand a two-second cycle and a sixty-second cycle quite distinctively. (Clayton, 2000)

This current study intends to include cultural factors into the discussion and exploration of the relationship between pulse, grouping and metrical structure, asking how this relationship could be influenced by one's previous experience of a particular musical culture. It also inquires whether pulse and meter are processed distinctively. If so, as suggested in the neuroscientific studies, how are the differences reflected in the behavioural response?

II. METHOD

Complex rhythm and long rhythmic cycles are commonly found in both classical Indian music and in the traditional Middle-Eastern music. In a behavioral experiment, ten loops of Turkish rhythmic patterns (representing the Middle-Eastern metrical structures) are presented to three groups of subjects that have listening experience with 1) Middle-Eastern music (metrically and culturally familiar group), 2) Indian music (metrically familiar but culturally unfamiliar group), or 3) no experience with music of complex meters (unfamiliar group). Among the ten rhythmic patterns, seven are of complex meter

(7-beat, 9-beat, and 10-beat); the other three are of simple meter (3-beat and 4-beat) as a control variable.

The experiment sessions contain two parts in which the participants are asked to tap to the acoustic stimuli. The first part tests the participant's entrainment to pulse (P), whereas the second part tests their tapping responses to the metrical or grouping structure of the patterns (M). In both parts all ten rhythmic patterns are played in two tempo versions (thus forty in total) — one in the range of the mean preferred tempo (P) of human's pulse perception that is around 500 ms, 120 bpm, and the other is in a faster temporal range where the pace is doubled (F). The temporal difference is used as a variable to manipulate the participants' choice of reference level when extracting a pulse and/or grouping. The results from the two temporal conditions are used to answer the question whether there is a disassociation in the response behaviour to the pulse and the meter task.

The order of the rhythmic pattern presentation is randomized in both the pulse and the meter tasks. Subjects' tapping responses and the sound stimuli are recorded simultaneously on separate, synchronized tracks for future analysis.

In addition to the behavioral responses, participants give verbal feedback about their entrainment strategy and fill out a brief questionnaire about their musical cultural background at the end of the experiment session.

III. RESULT AND DISCUSSION

The three preliminary analyses presented here are based on the responses from 24 subjects (n=24), with 8 from each cultural group: A (Middle-Eastern), I (Indian), and O (Others).

To assess the relationship between responses and experimental variables, the first analysis uses a generalized linear mixed model (GLM) with binomial error distribution and a log link function to compare the percentage of correct performance in the pulse task (P) and that of performance showing comprehension of the periodicity of the stimuli in the meter task (M), with *group*, *task*, *tempo*, and *complexity* (simple or complex meter) as fixed factors. A response counted as showing comprehension of the periodicity is when at least one tap per cycle was given at a fixed location within a pattern, without consideration of its metrical structure.

In this analysis, the main factor *group* is significant (p=0.003), with I having a larger difference than the other two (I>O>A) to indicate the importance of cultural influences. On the interactions between factors, except for the significance found between *task* and *complexity* (p<0.001), with Simple>Complex in the pulse task and Complex>Simple in the M (periodicity) task, no other significant interactions, such as that between *group* and *task*, are found. (Fig. 1) The result that group I has overall a better performance in all conditions is probably due to their advanced experience in complex metrical structures. Although the cultural issues are not addressed much in the current report, more detailed discussions regarding this phenomenon are worth elaborating in the continual research.

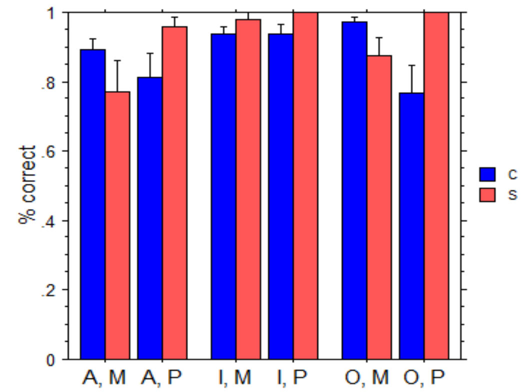


Figure 1. Analysis I: Task performance (mean % correct) for Pulse(P) vs. Periodicity(M). C=complex meter; S=simple meter; A=Middle Eastern; I=Indian; O=Other.

The second analysis is also based on the GLM using the same fixed factors, comparing more specifically the percentages of correct performance in the pulse task (P) and of the responses showing correct understanding of the metrical structure in the meter task (M). The result shows that the factors *task* (p<0.001) and *complexity* (p<0.001) are highly significant, but no significant interactions between the factors are found. Post-hoc analysis indicates that in all three cultural groups and both temporal conditions, better performance is achieved in the pulse task than in the meter task, and in the simple meters than in the complex meters. The tempo change has a large influence on the meter task and on the complex meters, but has only a limited effect on the pulse task and on the simple meters. The analysis suggests a disassociation between responses to pulse and to meter, as well as the possibly distinct mechanisms for complex and simple meters (Figs. 2 and 3). Additionally, there is a tendency for the A and I groups to perform better in the meter task for the fast tempo than for the slow tempo; this difference, however, did not turn out to be significant in the current study.

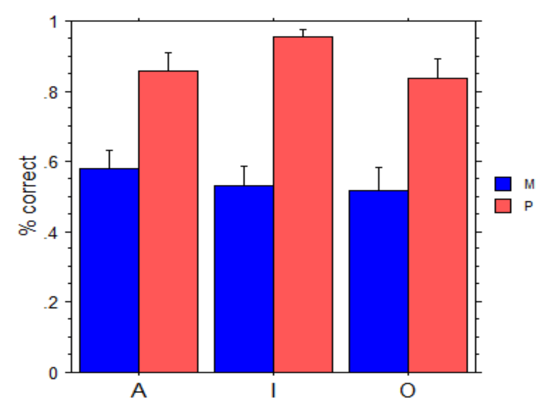


Figure 2. Analysis II: Task performance (% correct) Pulse(P) vs. Metrical perception (M). A=Middle Eastern; I=Indian; O=Other. Mean for AM: .581; AP: .856; IM: .531; IM: .956; OM: .519; OP: .838.

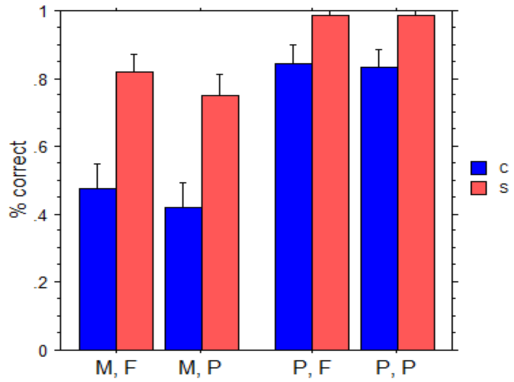


Figure 3. Analysis II: Task performance (% correct) Pulse(P) vs. Metrical perception (M). C=complex meter; S=simple meter; F=fast tempo; P=mean preferred tempo. Mean for MFC: .476; MFS: .819; MPC: .422; MPS: .750; PFC: .845; PFS: .986; PPC: .833; PPS: .986.

To look more closely at the relationship among the subjects' response, *task*, and *tempo*, the third analysis looks at how the overall tapping responses align with the stimuli for the pulse and meter tasks under the two temporal conditions. In the pulse task, the frequencies of the tapping levels (i.e. beat interval between taps) are compared. The result shows a clear shift in reference level under the two temporal conditions, from dominantly one-beat distance (tapping on every beat) at P (slower, mean preferred tempo), to two-beat distance (tapping on every other beat) at F (faster, doubled tempo; see Fig.4). However, in the meter task, no such level changes occur. Despite their larger variety, the overall tapping patterns and the agreement across subjects remain basically the same in the two temporal conditions. This suggests another disassociation of the response to the pulse and meter task — the tempo change leads to a shift in the reference level for pulse, but not for meter tapping.

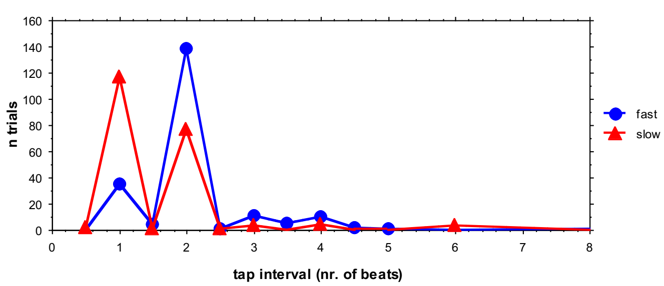


Figure 4. Analysis III: Tapping pattern to pulse under the two tempo conditions.

Also, there are no unambiguous "head nodes," as suggested in the dominant Western meter theories (Fitch, 2013). Examining the average tap frequencies in the meter task, it is found that in the slow (mean preferred temporal zone) 4-beat meter, beat 1 and 3 receive similar number of taps; in the 7-beat meter, beat 1 and 4 receive similar number of taps in both temporal conditions (Fig.5); beat 1 and 6 in the 10-beat patterns also receive similar tapping frequencies in both tempi. This analysis does not suggest any hierarchical processing in the meter task.

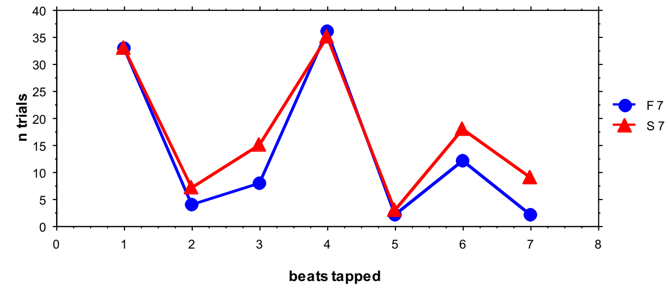


Figure 5. Analysis III: Tapping pattern to meter under the two tempo conditions. (Example of the 7-beat patterns)

IV. CONCLUSION

This experiment produced no evidence for hierarchical processing of musical meter. It found, instead, that participants showed different behaviours in response to the pulse and the meter task. The three preliminary analyses reported were able to distinguish between responses to the pulse, periodicity, and metrical aspects of the stimulus patterns. On the one hand, differences between responses to the pulse and the meter task (e.g. changing reference levels in the former) seem to suggest different underlying mechanisms between pulse and meter perception. On the other hand, differences in response to periodicity and metrical structure seem to be due to the complexity of the response pattern: there were no differences between simple and complex meter for tapping the periodicity, only for tapping the metrical structure. Further analyses would be necessary to test whether these differences are gradual or categorical. However, both response patterns can be explained on the basis of sequential processing of sound features of the stimulus patterns.

These results challenge Western hierarchical meter theories, and the preview of cultural influence on rhythmic perception raises awareness that the rhythmic understanding and practices of many cultures may not be on the same page with today's dominant theories.

ACKNOWLEDGMENT

The authors of this paper would like to thank the School of Music at the Ohio State University for all the support. G. B. G.

REFERENCES

Arom, S. (1991). *African Polyphony and Polyrhythm*. 1st ed. Cambridge, UK: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511518317

Clayton, M. (2000). *Time in indian music: Rhythm, metre, and form in north indian rag performance*. New York: Oxford University Press.

Cross, I. (2003). Music as biocultural phenomenon. *Annals of the New York Academy of Science, 999*, 106-111.

Fitch, W. T. (2013) Rhythmic cognition in humans and animals: distinguishing meter and pulse perception. *Frontiers in Systems Neuroscience* 7, 1-16.

Geiser, E., Zaehle, T., Jankcke, L. and Meyer, M. (2009) The Neural Correlate of Speech Rhythm as Evidenced by Metrical Speech Processing. *Journal of Cognitive Neuroscience*, Vol. 20, No. 3 , Pages 541-552.

Kuck H., Grossbach, M., Bangert, M. and Altenmueller, E. (2003) Brain Processing of Meter and Rhythm in Music. *Ann. N.Y. Acad. Sci,* 999: 244-253.

Will, U. (1999) La baguette magique d'ethnomusicologie. Repenser la notation et l'analyse de la musique.(The magic wand of Ethnomusicology. Re-thinking notation and its application in music analyses) In: Cahiers de Musiques Traditionelles, Vol. XII. Georg, Genève, pp.9-34.

Will, U., & Turow, G. (2011). Introduction to entrainment and cognitive ethnomusicology. In J. Berger, & G. Turow (Eds.), *Music, science, and the rhythmic brain: Cultural and clinical implications* (pp. 3-30). New York: Routledge.

Statistical learning for musical meter revisited: A corpus study of Malian percussion music

Justin London[1,*], Nori Jacoby[2,*], Rainer Polak[3,*]
[1]Department of Music, Carleton College, USA
[2]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, USA
[3]Institute for World Music, Cologne University of Music and Dance, Germany
[*]all authors contributed equally to this work

Background
Palmer & Krumhansl (1990) tallied note onsets in an corpus of European art music and found strong correlations between their metric distribution and normative patterns of accent. As metric accent may be read directly off of the cumulative distribution of note onsets, they posited a simple statistical learning process may play a role in the acquisition of metrical knowledge.

Aims
We wish to test Palmer & Krumhansl's influential hypothesis with music from a different musical culture whose characteristic rhythms are notable for their complexity, density and non-isochronous structure.

Method
15 recordings of 3 pieces of Jembe ensemble music (Manjanin, Maraka, and Woloso), all based on a 4-beat meter with ternary beat subdivision, were made in Bamako, Mali in 2006/07. Unidirectional microphones clipped to each drum cleanly captured the drumstrokes in each part. Sony Soundforge Pro and Steinberg Wavelab were used for onset detection, yielding some ~40,000 data-points with a temporal resolution of ±1ms for analysis.

Results
Metric positions here are indexed as x.y (x=beat & y=subdivision). The downbeat (1.1) aside, onbeat positions (2.1, 3.1, and 4.1) are articulated less frequently than the offbeat positions which precede them (x.3). Mid-beat positions (x.2) are articulated much less frequently than the surrounding positions, though a monotonic increase in frequency for the midbeat positions (from 1.2 to 4.2) is apparent. In Malian Jembe music metrical accent cannot be directly inferred from the statistical likelihood of a drumstroke's occurrence within the measure.

Conclusion
A simple statistical learning process cannot convey the metrical structure of Jembe music to its listeners. We propose that a modified form of statistical learning, supplemented by additional musical and extra-musical information, is required here; analogous cues may also play a role in Western musical enculturation.

Entrainment to an auditory rhythm: is attention involved?

Richard Kunert[1,2] and Suzanne R. Jongman[1]
[1]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[2]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands

Many natural auditory signals, including music and language, change periodically. The effect of such auditory rhythms on the human brain is unclear however. One widely held view, Dynamic Attending Theory, suggests that the general attentional system entrains to the rhythm and increases attention at moments of rhythmic importance. In support, two experiments reported here show reduced response times to visual letter strings shown at moments of auditory rhythmic importance, compared to moments of less importance. However, an attentional account would further predict rhythm entrainment to also influence memory for visual stimuli. However, in two pseudoword memory experiments we find evidence against this prediction. Whether a pseudoword is shown during a moment or auditory rhythmic importance or not is irrelevant for its later recognition memory in silence. These findings suggest that auditory rhythm perception is based on auditory-motor entrainment. Auditory rhythms are without an entraining effect on general attention.

Walking cadence is influenced by rhythmic regularity in speech

Simone Falk[1,3], Elena Maslow[2], Thierry Legou[3], Simone Dalla Bella[4,5,6], Eckart Rupp[1]
[1]Department of German Philology, LMU, Munich, Germany
[2]Institute of Phonetics and Speech Processing, LMU, Munich, Germany
[3]Laboratoire Parole et Langage, Aix-Marseille University, Aix-en-Provence, France
[4]EuroMov, University of Montpellier, Montpellier, France
[5]Institut Universitaire de France, Montpellier, France
[6]BRAMS, Montreal, Canada

In music, rhythmic structure helps to coordinate movement in time with the auditory stimulus. When we listen to music, we may snap our fingers or stomp our feet with the perceived beat of musical performances. Furthermore, moving to the beat impacts positively on the stabilization of physiological functions (e.g. breathing, heart rate) during activities such as cycling and walking. At the same time, moving to the beat affects the perception of musical rhythm and of the metrical structure of the stimulus. To date, it is unclear whether a mutual influence exists between rhythmic properties of speech and gross motor functions such as walking. In the present study, we investigate the link between rate variations of a regular speech rhythm and walking cadence. We recorded the walking cadence of 15 French-speaking listeners while they walked and listened to verbal stimuli. Stimuli were designed as a "verbal metronome" composed of lists of words (i.e., either nouns or verbs) that occurred at regular intervals. The verbal stimuli were presented at three different tempi (i.e., a base tempo +/-10%). Participants' preferential cadence was recorded before and after the experiment. We found that stimulus rate affected walking cadence. Participants slowed down when stimulus rate was slower than their preferential cadence. Moreover, step length increased when participants listened to lists of verbs compared to nouns. These results provide first evidence that perceived rhythmic properties of speech can influence gross motor functions such as walking. Moreover, we discuss the results on verb and noun semantics in light of recent embodiment theories.

Hierarchical Structure of Musical and Visual Meter in Cross-modal "Fit" Judgments

Stephen E. Palmer & Joshua Peterson

***Abstract*—**The metrical hierarchy of musical rhythm is defined by the structure of emphasis on beats in measures and has been studied in several ways [1]. Here we investigated the perceived structure of 3/4 and 4/4 time signatures in auditory and visual meter using cross-modal goodness-of-fit ratings for visual and auditory probes, respectively. In the *auditory context conditions*, four measures in 3/4 or 4/4 time were defined by a louder beat followed a series of 2 or 3 softer, equally-timed beats, respectively. A visual probe circle was introduced into the next four measures at one of 12 phase-angles relative to the auditory downbeat: 0, 45, 60, 90, 120, 135, 180, 225, 240, 270, 300, or 315 degrees. In the *visual context conditions*, context and probe modalities were reversed, with analogous visual rhythms being defined by a larger downbeat circle followed by a series of 2 or 3 smaller circles, with an auditory probe in the last four measures at one of the same 12 phase-angles. Participants rated how well the probe stimulus "fit" the rhythmic context in the other modality. For the visual context conditions, the probe's effect on fit-ratings revealed the expected beat-defined metrical hierarchy. In 4/4 time, fit ratings were highest for beat 1, next highest for probes at (or near) beats 2, 3, and 4, and lowest for probes at non-beats. In 3/4 time, they decreased similarly from beat 1 to beats 2 and 3, and from them to non-beats. The auditory context conditions produced unexpected results, however, with a single broad peak at and following the downbeat, and little evidence of elevated fit ratings for other beats over non-beats. Similar results were obtained when participants made explicit ratings of cross-modal synchrony using the same stimuli. Various factors relevant to explaining the asymmetry between these cross-modal conditions are discussed.

***Keywords*—Goodness-of-fit, Metrical hierarchy, rhythm, time signatures.**

I. Introduction

The metrical hierarchy of rhythm in music is defined by the structure of emphases related to beats in measures at multiple levels. This hierarchy is an important component of music theory because it structures the perceptual organization of music in time (e.g., Cooper & Meyer 1960; Lehrdal & Jackendoff, 1983) [2, 3]. Its empirical basis was explored most notably by C. Palmer and Krumhansl (1990) [1], who investigated three different measures of metrical structure: frequency distributions of note onsets in the scores of musical compositions, memory confusions in discrimination tasks, and goodness-of-fit judgments of temporal patterns in metrical contexts. In each case they found evidence of a multi-level metrical hierarchy of accent strength, with the strongest emphasis on the primary beat (downbeat) of each measure, and the next-strongest emphasis on the secondary beats of the time signatures they investigated: 2/4, 3/4, 4/4, and 6/8.

Most relevant to the issue of how listeners perceive temporal organization and structure in music, C. Palmer and Krumhansl (1990) [1] measured goodness-of-fit judgments in the auditory modality for higher-pitched probe beats relative to lower-pitched context beats. The context beats marked the downbeats of a series of several measures, and participants were asked to think of each context beat as the first of N beats, where N = 2, 3, 4, or 6. They were then asked to rate "how well the high-pitched (probe) beats fit with the (low-pitched) context beats" using a 7-point Likert scale, in which 7 indicated the best fit. The metrical hierarchy was evident in the pattern of their results: probe beats received higher ratings when they coincided with the primary and secondary beats of the implied metrical hierarchy than when they did not. These effects were stronger for people with more extensive musical training than for those with less training. Such experiments have restricted the probe and context modalities to auditory events, presumably because the temporal structure of music is primarily perceived in the auditory modality.

Nevertheless, the metrical structure of music and music-related events can also be evident in the visual, tactile, and kinesthetic modalities. Perhaps most obviously, the conductor of an orchestra or chorus moves her hands rhythmically to indicate metrical structure, typically using downward hand motions to indicate the primary beat of a measure (the downbeat) and sideways or upward motions to indicate secondary beats. These movements communicate metrical structure in the visual modality to musicians so that they can coordinate and synchronize their performances. Musicians themselves also move their hands and bodies rhythmically in time with the metrical hierarchy, as do novice and expert dancers alike. Further examples of aesthetic domains in which the timing of visual events provides information about the metrical structure of simultaneously played music include music videos and music visualizers.

In this report we describe our initial attempts to study the

Stephen E. Palmer is with the Department of Psychology, University of California, Berkeley CA 94720-1650 (phone: 510-684-4447; fax: 510-642-5293; e-mail: sepalmer@gmail.com).

Joshua Peterson is with the Department of Psychology, University of California, Berkeley CA 94720-1650 (e-mail: jpeterson@berkeley.edu).

Institute for Metals, Tsukuba, Japan (e-mail: author@nrim.go.jp).

cross-modal perception of musical meter in the visual and auditory modalities. We used an extension of C. Palmer and Krumhansl's (1990) [1] goodness-of-fit paradigm that we adapted for cross-modal stimuli by presenting larger and smaller white circles visually and playing louder and softer beats auditorily. Half of the participants rated the fit of a visual probe flash relative to an auditory context of loud and soft beats in 3/4 or 4/4 time, and the other half rated the fit of an auditory probe beat relative to a visual context of larger and smaller circles in 3/4 or 4/4 time. We expected that both cross-modal context conditions would essentially replicate the results of C. Palmer and Krumhansl (1990) [1], with the caveat that the auditory context might provide stronger evidence of metrical structure than the visual context because of its stronger associations with music perception. This did not turn out to be the case, however.

II. Methods

A. Design. The experimental design consisted of three orthogonal factors: *context condition* (visual or auditory), *duration condition* (long, medium, or short), and *probe phase-angle condition* relative to the downbeat (0°, 45°, 60°, 90°, 120°, 135°, 180°, 225°, 240°, 270°, 300°, or 315°). Context condition and duration condition were between-subjects variables, whereas probe phase-angle was a within-subject variable.

B. Participants. 24 undergraduates at the University of California, Berkeley, participated in the study. The first 8 were run in the long duration condition, the next 8 in the medium duration condition, and the last 8 in the short duration condition. In each duration condition, 4 participants ran in the visual context conditions and 4 in the auditory context conditions, randomly assigned across participants. Musical training varied unsystematically from 2 to 10 years, but we do not yet have enough participants in the sample to meaningfully examine training effects in comparable stimulus conditions.

C. Stimuli. In the auditory context conditions, four measures in 3/4 or 4/4 time were defined by a louder beat followed a series of 2 or 3 softer, equally-timed beats, respectively. A visual probe of a solid white circle was introduced into each of the next four measures at one of 12 phase-angles relative to the auditory downbeat: 0°, 45°, 60°, 90°, 120°, 135°, 180°, 225°, 240°, 270°, 300°, and 315°(see Fig. 1). In the visual context conditions, context and probe modalities were reversed, with analogous visual rhythms being defined by presenting a larger (downbeat) white circle followed by a series of 2 or 3 smaller white circles, with an auditory probe in the last four measures at one of the same 12 phase-angles.

The downbeats occurred at the 0° probe phase-angle in both the 3/4 and 4/4 time signatures, with each measure being 2 seconds long. The secondary beats occurred at 120° and 240° in the 3/4 time condition, and at 90°, 180°, and 270° in the 4/4 time condition as indicated in Fig. 1. The 60°, 180°, and 300° phase-angles are more closely related to the 3/4 condition because they are the "back-beats" or "half-beats" intermediate between the 3/4 beats at 0°, 120°, and 240°. In contrast, the 45°, 135°, 225°, and 315° phase-angles are more closely related to the 4/4 condition because they are likewise intermediate between the 4/4 beats at 0°, 90°, 180°, and 270°.

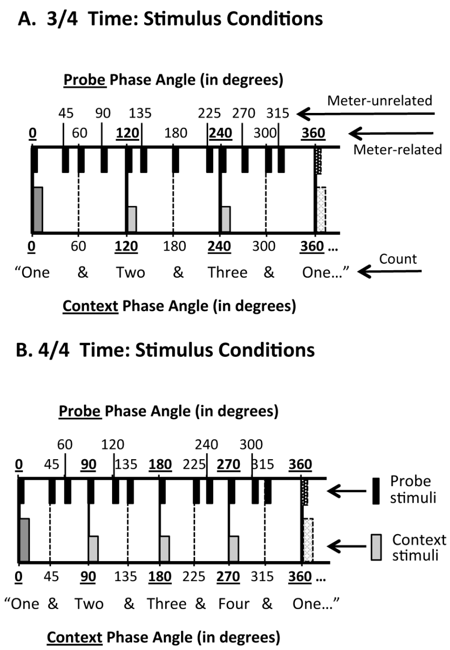


Fig. 1. Temporal structure of cross-modal stimulus events in 3/4 time (A) and 4/4 time (B). Context stimuli in one modality are represented in gray at the bottom, and probe stimuli in the other modality are represented at the top. Only one probe was presented per trial. Taller, darker context rectangles represent louder or brighter downbeat stimuli.

There were three event duration conditions, which will be referred to as long, medium, and short. In the long condition, the auditory stimuli were loud and soft snare drumbeats that lasted 250 ms including natural reverberation, and the visual stimuli were large and small circles that lasted 32 ms. In the medium condition, the snare drumbeat was trimmed to 100 ms (no reverberation and 50ms fade-out) and the visual stimuli were again 32ms. In the short condition, the auditory stimuli were loud and soft sinusoids at 523.25 Hz (C5) that lasted 16 ms (to match one 60 Hz monitor frame refresh cycle) and the visual stimuli were large and small circles that also lasted 16 ms (one frame). The long duration condition was studied initially to provide natural sounding auditory stimuli and natural looking visual stimuli. The medium and short duration conditions were included to test stimulus conditions with increasingly more precise timing to avoid cross-modal contamination in the synchronization of stimuli.

D. Procedure. Participants were asked to rate how well the probe stimulus (one event per measure) "fit" the rhythmic context of the events in the other modality (three or four beats per measure) using a 400-pixel horizontal line-mark scale. The ends were labeled "Good fit" and "Bad fit" with the left/right positions of the labels counterbalanced across subjects. Participants slid a restricted mouse cursor horizontally over the scale and clicked at the point they felt best represented how well the probe events fit the context events. Ratings ranged from -200 (worst fit) to +200 (best fit).

III. Results

The probe's position affected the fit-ratings quite differently for the visual and auditory context conditions (Fig. 2). When the context consisted of a temporal pattern of visual events and the probe was auditory, the results are reasonably similar to those of C. Palmer and Krumhansl (1990) [1], clearly showing the effects of the expected metrical hierarchy (Figs. 2A and 2B). The evidence for this claim comes from examining the differences between the fit ratings for the 3/4 and 4/4 time.

When the time signature of the visual contextual events was 3/4 (i.e., circles that were *large-small-small*, etc.), fit ratings were highest when the auditory probe coincided with the visual downbeat at 0° and next highest when it coincided with beats 2 and 3 (at 120° and 240°). Rating were also relatively high when the auditory probe was presented slightly after synchrony with visual beats 2 and 3 (at 135° and 270°),

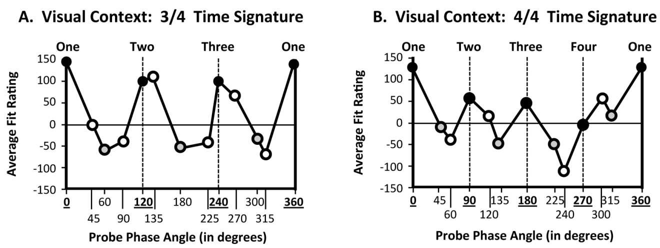


Fig. 2. Average fit ratings for auditory probes within visual contexts in 3/4 time (A) and 4/4 time (B).

presumably because they were relatively "close enough" to be perceived as coinciding with the visual event. There was essentially no evidence that fit ratings for the probes that coincided with the "back-beats" halfway between context beats (i.e., the gray circles in Fig. 2A at 60°, 180°, and 300°) were rated as fitting better than the probes that did not fit into the metrical hierarchy of the 3/4 time signature at all (i.e., the white circles in Fig. 2A at 45°, 90°, 225°, and 315°).

When the temporal time signature of the visual events was 4/4 (i.e., circles that were large-small-small-small, etc.), fit ratings were notably higher when the auditory probe coincided with the visual beats at 0°, 90°, 180°, and 270° than when they did not. Again, there was no indication that fit ratings for the probes that coincided with the "back-beats" midway between the beats (i.e., the gray circles in Figure 2A at 45°, 135°, 225°, and 315°) were rated as fitting better than the probes that did not fit into the metrical hierarchy of the 4/4 time signature (i.e., the white circles in Figure 2A at 60°, 120°, 240°, and 300°).

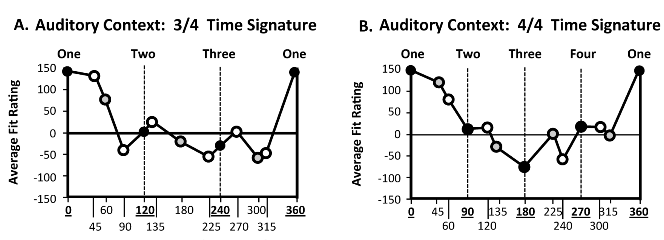


Fig. 3. Average fit ratings for visual probes within auditory contexts in 3/4 time (A) and 4/4 time (B).

To our surprise, the pattern of fit ratings was distinctly different when the context beats were auditory and the probe was visual (Fig. 3) than when the context beats were visual and the probe auditory (Fig. 2). In both the 3/4 (Fig. 3A) and 4/4 time signatures (Fig. 3B), there was a single broad peak around 0° that lasted through 45° and even 60°, dropping to near-zero only when the phase-angle reached 90°. This is in stark contrast with the results in the corresponding visual context conditions, where the auditory probes' fit ratings dropped to nearly zero by a 45° phase lag. We also note that the auditory contexts produced no clear differentiation between the structure of the fit ratings at later time lags in the 3/4 and 4/4 conditions, as were so obviously present for the visual contexts.

However, it is interesting to note that there are two quite modest peaks in the 3/4 condition (Fig. 3A) that occurred when the visual probes occurred just *after* the expected peaks at the 120° and 240° phase-angles, as if the visual probe was experienced a bit *after* the temporally synchronized auditory context events. The timing of these "late" peaks is also potentially consistent with the width of the peak for the downbeat that begins at 0°. That is, both effects are consistent with the possibility that the visual probes took longer to process and were temporally more variable when their fit was being evaluated relative to the auditory context events (i.e., when participants were attending to the visual probe as the to-be-judged event). The same tendency is not apparent in the 4/4 condition, however (Fig. 3B), perhaps because the more rapid series of auditory context events caused more extensive overlap.

Even so, it is interesting that any delay and/or variability in processing visual events does *not* seem to disrupt the perception of meter in the visual context condition where the probes are auditory. That is, peaks are clearly evident at the expected phase-angles in Fig. 2.

We decided to find out whether the asymmetries between the visual context and auditory context conditions in the cross-modal fit ratings might be due to corresponding asymmetries in people's ability to detect cross-modal simultaneity in these stimuli. We ran 8 additional participants asking them to rate "the alignment of the beep and flash in time" using the same rating scale. The results (Fig. 4) show the same pattern

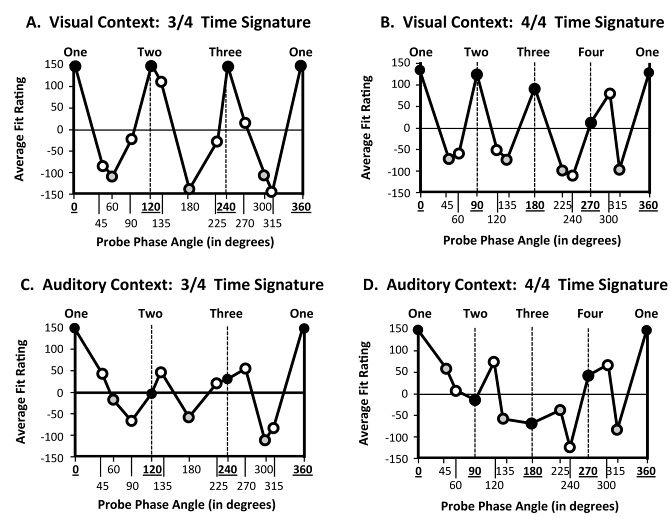


Fig. 4. Average synchrony ratings for visual contexts with auditory probes in 3/4 time (A) and 4/4 time (B) and for auditory contexts with visual probes in 3/4 time (C) and 4/4 time (D).

as the fit ratings. The synchrony ratings for the visual context conditions peak sharply at the synchronous cross-modal beats (Figs. 4A and 4B), and they are highly correlated with the corresponding fit ratings in Fig. 2 (r = .93, .84 for the 3/4 and 4/4 time signatures, respectively, p < .001). In contrast, the synchrony ratings for the auditory context conditions have sharp peaks only around the downbeat (near 0°), with reduced peaks just after the synchronous beats. They too are highly correlated with the corresponding fit ratings (r = .70, .80 for the 3/4 and 4/4 time signatures, respectively, p < .012).

IV. Discussion

Why might this asymmetrical pattern of cross-modal effects arise in both people's fit ratings and their simultaneity ratings? Several factors seem likely to be relevant.

First, there is a great deal of evidence from the sensory motor literature that visual processing is both slower and less accurate temporally than auditory processing (Repp, 2003 [4]; see Repp, 2005 [5], and Repp & Su, 2013 [6], for reviews). For example, people are able to synchronize finger-tapping at much faster beat-rates with auditory metronomes than visual ones. Slower and more variable processing of visual probes could explain several features of the present results: the broad peaks evident around the downbeat in the auditory context conditions (Figs. 3A and 3B), the greatly reduced secondary peaks around the on-beats in 3/4 time (Fig. 3A), and perhaps even the lack of secondary peaks around the on-beats in 4/4 time (Fig. 3B). Slower, more variable processing of visual events alone, however, fails to predict why the results from the visual context conditions are so much cleaner and more precise than those from the auditory context conditions (compare Figs. 2 and 3). If visual events cause problems only when they constitute probes and not when constitute the metrical context, some other factor(s) must be at work.

A second consideration is the differential role of temporal certainty/uncertainty in the context and probe events of the task itself. Participants initially perceive four measures of the context meter alone. This presumably allows them to create a precise temporal template for anticipating upcoming context beats. In measure 5, the first appearance of the probe is maximally uncertain, because it can occur anywhere in the measure. Its additional occurrences in measures 6-8 presumably reduce its temporal uncertainty, but not to the same degree that the context beats in measures 1-4 do. This consideration suggests that by the 5th measure, the visual context condition may have produced a very stable and accurate template of expectations for context beats, even if processing is slower and more variable in vision than audition.

Third, the two cross-modal conditions may not have been equally cross-modal, in that participants may well have converted the metrical visual context into quasi-auditory form by inwardly, or even outwardly, counting the meter throughout the trial ("***one***, two, three, ***one***, two, three," etc. for 3/4 time). This transformation would have the desirable effect of making the ostensibly cross-modal "visual context" condition much more nearly uni-modal, since both the probes and context would end up being represented auditorily.

Note that the same is unlikely to be true in the auditory context condition with the visual probe. If the optimal strategy is to convert the visual stimulus into an auditory one, perceivers would have to inwardly subvocalize to the visual probe. The problem is that the probe occurs at a very uncertain time.

It seems likely that any or all of three of these factors may be relevant to understanding both the fit ratings and the synchrony ratings. Further experiments manipulating variables related to these factors should help determine the extent to which they influenced the present results.

V. Conclusions

Cross-modal fit ratings visual contexts with auditory probes replicate and extend prior results on the metrical hierarchy for auditory probes in auditory contexts by C. Palmer and Krumhansl (1990). Corresponding ratings for visual probes relative to auditory contexts are quite different, however, showing elevated fit ratings primarily on and shortly after the louder auditory downbeat. Similar results were obtained when participants were explicitly asked to rate the degree of simultaneity of a probe within a context using the same stimuli. This correspondence strongly implies that people tend to find that synchronous cross-modal stimuli "fit well" together, and that difficulties in detecting cross-modal synchrony can lead to corresponding perceptions of poor fit. In future research we hope to track down the reason for these differences in different cross-modal combinations. C. Palmer and Krumhansl's (1990) prior results were for unimodal auditory-to-auditory stimuli – i.e., auditory probes to auditory contexts – using only downbeats as context. By systematically varying the probe and context modalities and the temporal nature of the contextual stimuli, we hope to be able to isolate the conditions under which this asymmetrical pattern of fit ratings is obtained and those under which it can be eliminated.

Acknowledgment

We thank Nori Jacoby, Caroline Palmer, and Carol Krumhansl for their comments on a previous draft of this paper. This research was sponsored in part by NSF Grant BCS-1059088 from the NSF to S.E.P.

References

[1] Palmer, C., & Krumhansl, C. L. (1990). Mental representations for musical meter. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(4), 728.

[2] Cooper, G., & Meyer, L. B. (1960). The rhythmic structure of music. University of Chicago Press, Chicago.

[3] Lerdahl, F., & Jackendoff, R. (1983). An overview of hierarchical structure in music. *Music Perception: An Interdisciplinary Journal*, *1*(2), 229-252.

[4] Repp, B. H. (2003) Rate Limits in Sensorimotor Synchronization With Auditory and Visual Sequences: The Synchronization Threshold and the Benefits and Costs of Interval Subdivision, Journal of Motor Behavior, 35:4, 355-370.

[5] Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic bulletin & review,* 12(6), 969-992.

[6] Repp, B. H., & Su, Y. (2013). Sensorimotor synchronization: a review of recent research (2006–2012). *Psychonomic Bulletin & Review,* 20(3), 403-452.

Musicianship and Tone-language Experience Differentially Influence Pitch-duration Interaction

Yu Fan,*[1] Udo Will*[2]

[*] *Cognitive Ethnomusicology, The Ohio State University, Columbus, OH, USA*
[1]fan.264@osu.edu,[2]will.51@osu.edu

ABSTRACT

Our understanding of rhythm perception has been influenced by studies that show an interaction of pitch and duration perception for both speech and non-speech sounds. Different interaction effects have been found for filled and empty intervals: filled higher pitched intervals are perceived as longer than lower pitched ones, while empty intervals marked by lower pitched tones are perceived as longer than those marked by higher pitched tones. Though language background has been shown to affect the degree of interaction for filled intervals, it is currently not known whether language and music experience influence the interaction for empty intervals. As both tonal language speakers and musicians show improved pitch processing capabilities, the current study examines the influence of language and music background on the pitch/duration interaction for empty intervals. 56 participants (Mandarin or English speaking musicians and non-musicians) performed an ordinal comparison test on pairs of empty intervals. Stimuli consisted of a reference interval (either 300 or 600 ms) followed by a target interval that varied as a percentage of the standard interval. The pitch of target tones were either the same as, or higher or lower than that of standard tones. Participants judged whether the target intervals "speed up" or "slow down" compared with the standard interval. Results show that duration perception in tone-language speakers is less effected by pitch than in non-tone-language speakers and less in musicians than in non-musicians. Interestingly, female non-musicians showed a greater effect of pitch-duration interaction than male non-musicians, but no gender effect was found for musicians. Our experiment demonstrates that both tone-language and music training can influence timing behavior by affecting the interaction between pitch and duration perception of empty intervals.

I. INTRODUCTION

Both pitch and timing are important properties of language and music that can influence human perception of emotion and meaning. In tone languages pitch variation is associated with the specific lexical meaning of a syllable (Wang & Sun, 2015). In music, the pitch and timing are important properties for expressing and inducing emotions (Juslin, 2005), and variations in pitch and timing can help the listener to judge the emotional state of a speaker (Belin, 2006). Various studies on human auditory processing have shown that perception of duration can be modulated by pitch changes (Brigner, 1988; Boltz, 1998; Henry & McAuley, 2009). Two different types of duration intervals have been described: filled intervals, in which an auditory signal lasts from the onset to the offset, and empty intervals, for which the duration equals the interval between two successive brief sounds. Studies on the interaction of pitch and filled intervals (Brigner, 1988; Matthews, 2013) have shown that the duration of higher pitched tones are judged longer than that of lower pitched tones. However, inverse results have been reported for the interaction of pitch and empty intervals: Pfeuty & Peretz (2010) found that the duration of empty target intervals is perceived as shorter when bounded by high tones, and as longer when bounded by low tones. Lake and coworkers (2014) confirmed these results, and offered an explanation of this time distortion within the framework of scalar timing theory.

Studies have shown that speaking a tonal language (Krishnan et, al., 2005) or being a musician (Wong et, al., 2007) improves pitch processing at the brainstem level. Consequently, various studies have explored links between language or music background and pitch-duration interaction. A number of linguistic studies have found ubiquitous pitch-duration interaction across languages (Gussenhoven & Zhou, 2013; Šimko et, al., 2015). For instance, high level tones (level tone with high pitch) occur in words containing short vowels while low level tones (level tone with low pitch) occur in words with long vowels, such as in Hu, a Mon-Khmer language (Jan-Olaf, 1988), Thai (Gandour, 1977), Cantonese (Kong, 1987), Bai, Tibetan, Zulu, etc (Faytak & Yu, 2011). All stimuli tested in these studies were real vowel sounds and hence filled interval stimuli. So far studies have not explored the interaction of pitch and empty intervals in linguistic context.

Other studies have shown that musicians are better at processing of pitch information (i.e. distinguishing and memorizing) and show a greater sensitivity in duration discimintion tasks than non-musicians (Gaab & Schlaug, 2003; Besson et al., 2007; Jeon & Fricke, 1997). Similar to research on languages, all these studies tested filled intervals. Differences in empty interval perception between musicians and non-musicians have not been considered yet.

The aim of present study is to explore the effects of cultural factors like language and music background on the interaction between pitch and interval duration perception. Specifically, we tested whether tonal language experience (Native Mandarin speakers) and professional musical training (Musicians) lead to less influence of pitch related distortion on timing estimation. We also explored whether the effect of these culture factors were different for shorter and longer inter-stimulus intervals. To achive these goals we tested four groups of participants: Native Mandarin-speaking musicians and non-musicians, Native English-speaking musicians and non-musicians. They were asked to complete an ordinal comparison task in two ISI duration condition (300 ms/600 ms). In the task, they need to compare whether or not the Target Intervals (TIs) sounded like "speed up" or "slow down" compare to the preceding standard ISIs.

II. EXPERIMENT

This experiment examined how language background and musical training affect pitch/duration interaction in empty interval perception. We hypothesize that duration perception is less influenced by pitch in tonal language speakers (e.g. Mandarin speakers) than non-tonal language speakers (e.g. English). We also hypothesize that it is less influenced in musicians than in non-musicians. If our experimental results support these two hypotheses, it would be a further indication that cultural factors like tonal language background and music training influence timing behavior. On the basis of previous studies, we also predict that the pitch/duration interaction is temporally limited, i.e. will be stronger for shorter inter-stimulus intervals than for longer ones.

1. Participants

For this experiment, we recruited fifty-six healthy adults who were either undergraduate or graduate students in the Ohio State University-Columbus. Both Mandarin-speaking musicians (hereafter referred to as MM) (n = 14; 7 males) and English-speaking musicians (hereafter referred to as EM) (n = 14; 7 males) were professional instrumentalists, vocalists or conductors with at least 8 years of continuous training in Western classical music. Mandarin-speaking non-musicians (hereafter referred to as NNM) (n = 14; 7 males) were born in mainland China and reported using their native Mandarin throughout their daily activity, though they learnt English as their second language since in high school. None of native English-speaking non-musicians (hereafter referred to as ENM) (n = 14; 7 males) had any experience with Mandarin. Moreover, both MNM and ENM had had no formal music training. These four groups were closely matched in age (MM: M = 26.07; SD = 3.03, EM: M = 24.14; SD = 4.12, MNM: M = 26.77; SD = 1.67; ENM: M = 24.31; SD = 3.12), and all participants reported normal vision and normal hearing. All participants gave their verbal consent in compliance with the study protocol approved by the Ohio State University Institutional Review Broad, and received a monetary compensation for their participation.

2. Stimuli and Apparatus

Stimuli were constructed as in the Lake et al. (2014) study. All tone waveform are created by MATLAB software. They were 50ms in duration with a 3ms linearly ramped rise and fall time, and a 44,100Hz sampling rate. In order to facilitate distinction of standard tones from target tones, the former had a fundamental frequency of 500Hz and two added higher harmonics (1000 and 1500Hz with 7:2:1 amplitude ratios). Stimuli were presented by MATLAB software, and delivered dichotically via headphones.

3. Procedure

Participants were tested individually in the Ethnomusicology lab at the Ohio State University-Columbus. Before the start of the experiment, participants completed a questionnaire concerning their linguistic and musical background. The whole experiment was presented by MATLAB software. After hearing a trial of tones via headphone, participants were instructed to judge "whether the Target Intervals (Tis) sounded sped up or slowed down compared to the standard ISIs." They had 3000ms to make their judgment by tapping either the "speed up" or the "slow down" button on the screen.

In each trial, participants heard four tones. The first and second tones, the standard interval, both had a fundamental frequency of 500 Hz. The third and fourth tone either had the same (500 Hz), a higher (625 Hz), or a lower (400 Hz) frequency. The inter-stimulus interval (ISI) between the first three tones was either 300ms or 600ms. For the shorter ISIs, the TIs were 240, 264, 284, 316, 336 or 360ms, for the longer ISIs, they were 480, 528, 576, 624, 672 or 720ms. Each combination of TI pitch, ISI length and TI length was repeated 9 times in one test. As a result, the total number of trials in one test was $3\times2\times6\times9=324$. The test was divided into 2 sessions of 162 trials according to ISI length (300/600ms). Trials in each session were randomly selected and grouped into 3 runs of 54 trials. Participants were given the opportunity to rest between runs and they did not receive any feedback about their performance.

III. RESULTS AND DISCUSSION

Two participants (one MNM and one ENM) had to be excluded from analysis because of inadequate task performance. Therefore the final sample size was fifty-four (27 males, 27 females; 18-30 years old; M = 25.31; SD = 3.33). For each participant, the proportion of "long" responses, P ("long" response), was computed as a function of TI length conditioned on ISI length (300/600ms) and TI tones (higher, same or lower than ISI tones). Fig. 1a &b show the averages for the three TI pitch conditions, and the fitted psychometric functions for each language/music background group.

Though we explicitly instructed participants to try not to miss any trial, there were a number of no-response trials for all participant groups (Table. 1). A chi-square test of independence was performed to examine the relation between no-response trials and the four participant groups. The result shows that there is a highly significant unequal distribution of no-response trails across the groups, $\chi^2(1, N = 119) = 10.8999, p = 0.00096 < 0.001$. Besides, Chinese musicians show the least number of non-responses, but, the Chinese non-musicians show the highest number.

Table 1. No-response trials distribution among each group (MM = Mandarin-speaking Musician; MN = Mandarin-speaking Non-musicians; EM = English-speaking Musician; EN = English-speaking Non-musician).

Group	Number of No-response Trials
MM	14
MN	45
EM	33
EN	27

Point of Subjective Equality (PSE) scores and ANOVA

The point of subjective equality (PSE) is the point at which one variable quantity is perceived equal to fix valued quantity by test subjects. In this study, PSE is the point when subject perceives the Target Interval (TI) as equal to the former standard ISI with fixed length. Particularly, PSE score for each FI tones is used to evaluate the effect of pitch level on

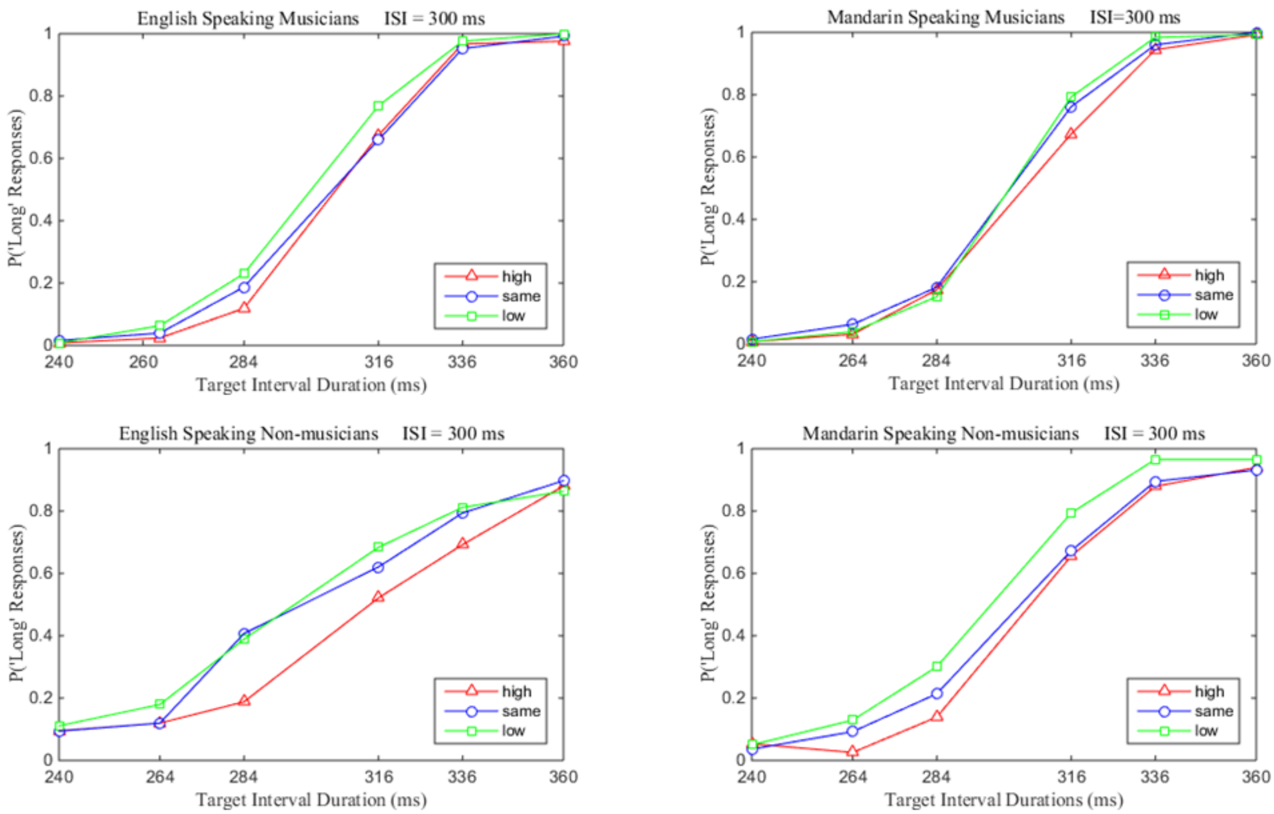


Figure 1a. **Group averages of the proportion of "long" responses on 300ms ISI grouped by participants' language background and professional musical training.**

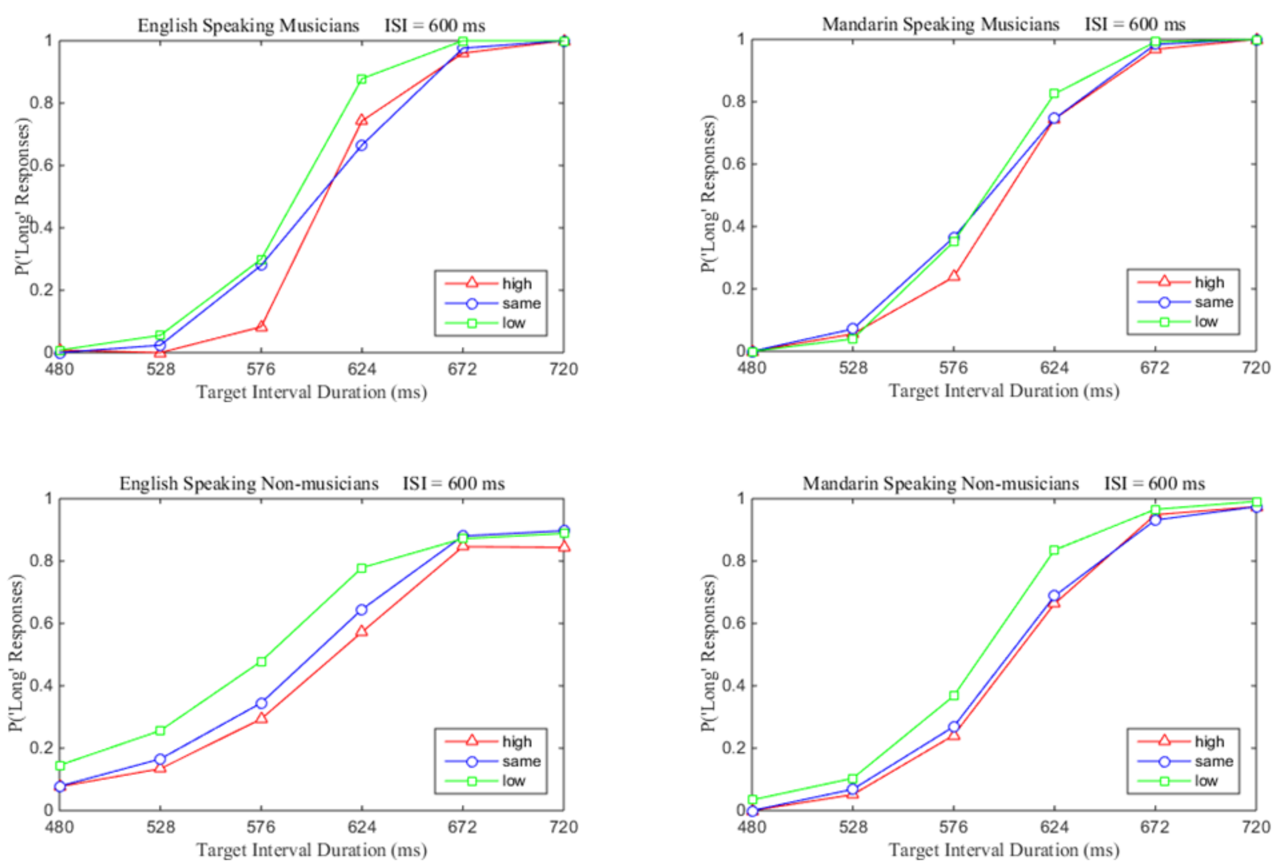


Figure 1b. **Group averages of the proportion of "long" responses for 600ms ISI grouped by participants' language background and professional musical training.**

Table 2. The values in the table indicate the mean PSE score of each group in three TI (Target Interval) tones, and the values in the parenthesis indicate its standard deviation.

Group	Standard ISI Interval 300 ms			Standard ISI Interval 600 ms		
	TI Tones			TI Tones		
	High	Same	low	High	Same	Low
MN	309.47(14.85)	304.1(12.05)	294.06(11.37)	606.16(21.41)	603.41(24.92)	585.79(23.98)
EN	315.86(13.83)	305.32(12.15)	297.6(17.74)	615.57(19.32)	602.77(15.03)	577.95(24.63)
MM	304.20(9.33)	300.47(15.03)	300.27(12.9)	599.41(17.1)	592.43(13.76)	590.16(17.02)
EM	303.2(11.93)	301.43(10.37)	298.67(10.34)	611.61(11.22)	602.7(15.4)	588.67(12.76)

the time perception of test subjects with different backgrounds.

We used quickpsy package in *R* to fit a psychometric function to our data. In the function formula (see below, with α = threshold, β = slope, λ =lapse rate), we fixed γ (guess rate) to 0, threshold and slope were set as free parameters, and the lapse rate limited to a range of $[0, 0.1]$. The PSE score is defined as the value of which $f(PSE) = 0.5$. The mean value of each group's PSE score is shown in Table. 2.

$$P_{response=long} = f(TI) = \gamma + (1 - \gamma - \lambda) \cdot \frac{1}{1 + e^{-\left(\frac{TI-\alpha}{\beta}\right)}}$$

Fig.1a and b indicate that target intervals tend to be perceived as shorter when they have a higher frequency than the standard interval, and as longer when their frequency is lower. In order to compare the PSE scores across the ISI condition we normalized them according to the equation: [PSE (High) – PSE (Low)]/PSE (Same). The normalized PSE scores were then subjected to a 2×2×2 mixed model ANOVA with language (Mandarin/English), music background (Musician/Non-musician), gender (Male/Female) as between-subject factors, and ISI duration as within-subject factor.

The results show significant main effects for language (F (1, 52) = 7.12, p = 0,010, Cohen's d: 0.726), music training (F (1, 52) = 12.29, p = 0.001, Cohen's d: 0.955), and gender (F (1, 52) = 6.74, p = 0.013, Cohen's d: 0.706). Post-hoc analyses show that duration perception in tone language speakers is less effected by pitch than in non-tone language speakers, in musicians less than in non-musicians, and in male participants less than in female participants. There is a significant interaction between music training and gender: F (1, 52) = 4.38, p = 0.042, Cohen's d = 0.531. In both Mandarin and English speakers, female non-musicians show a greater effect of the pitch-duration interaction than male non-musicians, whereas there is no significant gender difference for musicians. The effect of ISI duration was: F (1, 52) = 3.81, p = 0.057. As our prediction for this effect was directional, the appropriate test is one-tailed and the result is significant at α = 0.5. This means that for the longer ISI duration (600ms) the effect of the pitch-duration interaction is smaller than for the shorter ISI duration (300ms).

CONCLUSION

This study examined the effects of language and music background on the interaction between duration and pitch perception for empty intervals. In agreement with previous studies, our results show that if target intervals are higher pitched than standard intervals then their duration is underestimated, and it is overestimated if they are lower pitched. In addition, they demonstrate that for tone language speakers and professional musicians the interaction effect is significantly reduced. This suggests that long-term practices requiring extraction and evaluation of pitch information changes the interaction between pitch and duration processing. Unexpectedly, we also found a significant interaction between musical training and gender. The cause of this interaction is currently not known, but the effect merits further investigations that could improve our understanding of the mechanisms underlying the pitch-duration interaction.

REFERENCES

Belin, P. (2006, December 29). Voice Processing in Human and Non-human Primates. *Philos Trans R Soc Lond B Biol Sci, 361*(1476), pp. 2091-2107.

Besson, M., Schön, D., Moreno, S., Santos, A., & Magne, C. (2007, June). Influence of Musical Expertise and Musical Training on Pitch Processing in Music and Language. *Restorative Neurology and Neuroscience, 25*(34), 399-410.

Boltz, M. G. (1998, November). Tempo Discrimination of Musical Patterns: Effects Due to Pitch and Rhythmic Structure. *Perception & Psychophysics, 60*(8), 1357-1373.

Brigner, W. L. (1988). Perceived Duration as a Function of Pitch. *Perceptual and Motor Skills, 67*, 301-302.

Faytak, M., & Yu, A. C. (2011). A Typological Study of the Interaction between Level Tones and Duration. In E. Zee (Ed.), *Proceedings of the International Congress of the Phonetic Sciences XVII*. International Congress of the Phonetic Sciences.

Gaab, N., & Schlaug, G. (2003, December). The Effect of Musicianship on Pitch Memory in Performance Mached Groups. *NeroReport, 14*(18), 2291-2295.

Gadour, J. (1977). On the Interaction Between Tone and Vowel Length: Evidence from Thai Dialects. *Phonetica, 34*, 54-65.

Gussenhoven, C., & Zhou, W. (2013). Revisiting Pitch Slop and Height Effects on Perceived Duration. *14th Annual Conference of the International Speech Communication Association.* Lyon, France.

Henry, M. J., & McAuley, D. J. (2009, March). Evaluation of an

Imputed Pitch Velocity Model of the Auditory Kappa Effect. *Journal of Experimental Psychology: Human Perception and Performance, 35*(2), 551-564.

Jan-Olaf, S. (1988). Hu-a Language with Unorthodox Tonogenesis. (J. Davidson, Ed.) London: SOAS: Contributions to Mon-khmer Studies: Essays in honor of Professor H.L.Shorto.

Jeon, J. Y., & Fricke, F. R. (1997, April). Duration of Perceived and Performed Sounds. *Psychology of Music, 25*(1), 70-83.

Juslin, P. N. (2005). From Mimesis to Catharsis: Expression, Perception, and Induction of Emotion in Music. In D. Miell, R. MacDonald, & D. J. Hargreaves, *Musical Communication.* Oxford University Press.

Pfeuty, M., & Peretz, I. (2010, April). Abnormal Pitch-Time Interference in Congenital Amusia: Evidence from an Implicit Test. *Attention, Perception, & Psychophysics, 72*(3), 763-774.

Šimko, J., Aalto, D., Lippus, P., Wlodarczak, M., & Vainio, M. (2015). Pitch, Perceived Duration and Auditory Biases: Comparison Among Languages. Proceddings in ICPHS.

Kong, Q.-M. (1987). Influence of Tones upon Vowel Duration in Cantonese. *Language and Speech, 30*, 387-399.

Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. (2005). Encoding of Pitch in the Human Brainstem is Sensitive to Language Experience. *Cognitive Brain Research, 25*(1), 161-8.

Lake, J. I., LaBar, K. S., & Meck, W. H. (2014). Hear It Playing Low and Slow: How Pitch Level Differentially Influences Time Perception. *Acta Psychologica, 149*, 169-177.

Matthews, W. J. (2013). How Does Sequence Structure Affect the Judgement of Time? Exploring a Weighted Sum of Segments Model. *Cognitive Psychology, 66*, 259-282.

Wang, W. S.-Y., & Sun, C. (2015). *The Oxford Handbook of Chinese Linguistic.* Oxford University Press.

Wong, P. C., Skoe, E., Russo-Ponsaran, N., Dees, T., & Nina, K. (2007, May). Musical Experience Shapes Human Brainstem Encoding of Linguistic Pitch Patterns. *Nature Neuroscience, 10*(4), 420-422.

The Song Remains the Same: Biases in Musical Judgements and the Illusion of Hearing Different Songs

Manuel Anglada-Tort[1] and Daniel Müllensiefen[1]
[1]Department of Psychology, Goldsmiths, University of London, London, UK

The heuristics-and-biases framework developed by Tversky and Kahneman challenges the assumption of human rationality and is the fundamental theory underlying behavioural economics. However, it has not yet been applied explicitly to the field of musical judgements and choice behaviour. This study aims to answer to what extent people use rational judgements when evaluating music, and to what degree Kahneman's framework may be useful for the study of musical judgements and preferences. In order to address these questions, we made use of *the illusion of hearing different songs* and investigated how judgement biases affect participants' evaluations of music. Specifically, we tested the effects of mere-exposure and contextual information. 72 participants were misled to think that they had heard three different performances of the same musical piece, when in fact the exact same recording was played three times. Each time the recording was accompanied by a different text inducing low, medium or high prestige of the performer. We repeated this procedure with two different pieces of music, an excerpt from a classical piece that was generally unfamiliar to participants, and a piece of popular music that was more familiar. Most participants (77%) were prone to the deception, believing they had heard different musical performances. Participants' judgements showed a clear mere-exposure and contextual information effect. Musicians were not any more or any less susceptible to the illusion than non-musicians. Also, the results for both musical pieces were very similar. In conclusion, this research shows that when individuals attempt to evaluate music they rely on cognitive biases and heuristics that do not depend on the music itself, giving rise to 'non-rational' outcomes. Thus, the field of musical judgements and preferences would profoundly benefit from theories that take into account the non-rational character of human choice behaviour, such as Tversky and Kahneman's framework.

Being Moved by Unfamiliar Sad Music is Associated with Empathy

Tuomas Eerola[1,2] and Jonna Vuoskoski[1,3]
[1]Department of Music, University of Jyväskylä, Jyväskylä, Finland
[2]Department of Music, Durham University, Durham, UK
[3]Department of Music, Oxford University, Oxford, UK

The paradox of enjoying listening to music that evokes sadness is yet to be fully understood. Unlike prior studies that have explored potential explanations related to lyrics, memories, and mood regulation, we investigated the types of emotions induced by unfamiliar, instrumental sad music, and whether these responses are consistently associated with individual differences.

102 participants were drawn from a representative sample to minimize self-selection bias. Participants listened to an unfamiliar instrumental sad music, and psychophysiological responses (EDA and HRV) were recorded during the listening, and self-reports and indirect measures of emotion were collected immediately after the listening task.

The results suggest that the emotional responses induced by unfamiliar sad music could be characterized in terms of three underlying factors: Relaxing sadness, Moving sadness, and Nervous sadness. Moving sadness captured an intense experience that involved feelings of sadness and being moved, while Relaxing sadness was characterized by felt and perceived peacefulness and positive valence. Nervous sadness was associated with felt anxiety, perceived scariness and negative valence. These interpretations were supported by indirect measures of felt emotion and physiology. Experiences of Moving sadness were strongly associated with high trait empathy and emotional contagion, but not with other previously suggested traits such as absorption or nostalgia-proneness. Relaxing sadness and Nervous sadness were not significantly predicted by any of the individual difference variables.

The findings are discussed within a theoretical framework of embodied emotions and contrasted with the recent music and sadness studies using self-selected music. The implications about the specific mechanisms responsible for the experiences (e.g., episodic memories vs. emotional contagion) will also be considered.

Situational and dispositional influences on the functions of music listening

Fabian Greb[1], Wolff Schlotz[1], Jochen Steffens[1,2]
[1]Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, Germany
[2]Audio Communication Group, Technische Universität Berlin, Germany

Research on the functions of music listening mainly conceptualizes the functions as traits, whereas the potential variability across situations receives scant attention. Although many researchers mention the importance of situational influences on the way we interact with music, research on situation sill is in its infancy. Hence, this study aimed to differentiate between dispositional and situational influences on the functions of music listening and to reveal their relative importance. Another goal was to identify the most important situational and dispositional variables predicting the functions of music listening. To this end 587 persons completed an online study. Each participant sequentially described three self-selected listening situations and reported on situational characteristics (e.g. presence of other people, mood), the functions of music listening, and the music they usually listen to in the specific situation. After describing the listening situations, participants reported on sociodemographics and traits formerly shown to correlate with the functions of music (e.g. Big Five, musical taste). Mixed model analyses revealed that on average 36% of the variance of the functions was due to differences between persons and 64% of the variance was attributable to within person differences between situations. For further analyses all situational variables were within-subject-centered to separate situation-related from person-related effects. Several situational predictors such as activity or presence of other people were shown to have significant effects on all functions. In conclusion, the study gives valuable insight into how situational factors affect the functions of music listening. It further supports the conceptualization of the functions of music as both state and trait. As this study is part of a project aiming to predict music selection behavior, the potential significance of the functions as predictors will be analyzed in further steps.

Affording social interaction and collaboration in musical joint action

Ashley E. Walton[1], Auriel Washburn[1], Elaine Hollensbe[2], Michael J. Richardson[1]
[1]Department of Psychology, University of Cincinnati, Cincinnati, OH, USA
[2]Department of Management, University of Cincinnati, Cincinnati, OH, USA

Many daily activities require the coordination of actions with others, including navigating a crowded sidewalk, dancing with a partner, and participating in musical groups. Like everyday social behavior, musical joint action emerges from the complex interaction of environmental and informational constraints, including those of the instruments and the performance context. Music improvisation in particular is more like everyday interaction in that dynamics emerge spontaneously without a rehearsed score or script. Here we examined how the structure of the musical context affords and shapes interactions between improvising musicians. Six pairs of professional piano players improvised with three different backing tracks while their movements were recorded using a wireless motion tracking system. Each backing track varied in rhythmic and harmonic information, ranging from a chord progression, to a single tone. Afterward, while viewing videos from the performance trials, musicians narrated how they decided what to play and when. Narratives were analyzed using grounded theory (a qualitative method) to identify themes. For backing tracks with more structure, themes included *expertise* and *signaling*; for backing tracks with less structure: *personality, novelty,* and *freedom*. Differences in movement coordination and playing behavior were evaluated using linear and non-linear time series methods, to provide an understanding of the multi-scale dynamics that create the potential for musical collaboration and creativity. Collectively, our findings indicate that each backing track afforded the emergence of different coordination dynamics with respect to how they played together, how they moved together, as well as their experience collaborating with each other. Musical improvisation therefore provides a way to understand how social interaction emerges from the structure of the behavioral context, and how this structure supports coordination and collaboration in everyday behavior.

Empathy promotes interpersonal coordination through music

Peter Keller[1], Zoe Mitsopoulos[1], and Giacomo Novembre[1]
[1]The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Sydney, Australia

Coordinating one's actions with others in musical contexts requires sensorimotor and cognitive skills that may also support more general social interaction. Previous research has revealed relationships between individual differences in these skills and personality traits, including empathy, in tasks requiring musicians to synchronize movements with computer-controlled auditory sequences. We investigated whether empathy improves real-time interpersonal coordination between non-musicians engaged in musical activity and tested whether these effects generalize across situations with different temporal prediction demands, operationalized by varying musical leadership. Following pre-assessments of empathy, 13 pairs of participants with high perspective-taking scores and 13 pairs with low perspective-taking scores performed a musical coordination task involving the production of a familiar melody in synchrony on two electronic music boxes. Rotating the music box handles controlled the tempo, triggering piano tones separated by two octaves between co-performers. Individuals took turns at leading the interaction in half of the trials and, in the other half (order randomised), there was no assigned leader. Results indicate that the degree of asynchrony between tones produced by members of high-empathy pairs was smaller than in low-empathy pairs. Furthermore, while interpersonal synchrony was generally poor when leadership was assigned, high- and low-empathy pairs did not differ in this impairment. Findings suggest that empathy facilitates interpersonal synchrony in individuals without musical training. General perspective-taking abilities may enable individuals to predict the timing of other's actions via a process of motor simulation honed through everyday social interaction. The generalizability of this benefit across conditions where one or both individuals bear responsibility for predicting the other's action timing suggests that such perspective taking is automatic.

Creating Together: Improvisation in music can enhance social bonding

Katie Rose Sanfilippo[1], Jacques Launay[2], Eiluned Pearce[2] and Lauren Stewart[1]
[1] Department of Psychology, Goldsmiths College, University of London, London, UK
[2] Department of Experimental Psychology, Oxford University, Oxford, UK

Music has the ability to create social bonds and unify communities. Previous research has attributed this effect to the synchronization of movement between individuals, mediated by the release of endorphins. However, improvisation may also be an important contributor to this social bonding effect. Group improvisation involves creativity and communication, and comes with the shared reward of completing a complex task. This study aims to investigate whether improvisation can enhance social bonding when controlling for the effects of synchronisation. Fifty-five participants were assigned to one of three conditions: *unison* (complete synchrony, no improvisation, low complexity), *4-part harmony* (partial synchrony, no improvisation, moderate complexity), and *improvisation* (partial synchrony, full improvisation, high complexity). Social bonding was measured using a self-report questionnaire and a pre-/post- pain threshold test (as an index of endorphin release). Self-reported social bonding differed marginally between all three conditions ($p = 0.08$): bonding was significantly higher for *improvisation* than for *4-part harmony* ($p = 0.007$), but there were no significant differences between the other conditions. These results suggest that the effects of music-making on social bonding may be complex, involving several different factors such as reward and creativity. Both synchronization (unison) and improvisation seem to foster stronger social bonds that 4-part harmony. Creativity in group behaviour has an important role to play in fostering social ties. Its connection with social bonding helps explain why real musical activities might be more important in creating community connections than just tapping in time. These outcomes contribute to a better understanding of the evolutionary purpose of music and may have implications for how arts interventions are structured and used in clinical and educational settings.

Synchronous movement enhances cooperation in young children

Tal-Chen Rabinowitch[1] and Andrew Meltzoff[1]
[1]Institute for Learning and Brain Sciences, University of Washington, Seattle, WA, USA

Playing music together is a highly cooperative joint activity, requiring precise coordination between individual players. Certain features inherent to the performance of music may facilitate the intricate cooperation that enables players to create coherent music together. One such prominent feature is synchronization, the maintaining and alignment of rhythm between players. We wished to examine whether inter-personal synchrony is, in general, sufficient for enhancing cooperation, and whether this effect requires long-term experience with music performance or is rather present already early in development. To this end we designed an experiment intended for young children that decouples synchronization from cooperation. We swung four-year-olds on a swing set either in synchrony or asynchrony and subsequently tested how well they performed cooperative tasks. We found that brief synchronous swinging was sufficient to prime the children and significantly improve their cooperation compared to asynchrony or no swinging at all. We further revealed that synchronization increased their tendency to signal to each other more clearly their intended movements. These results demonstrate that by enhancing intention signaling synchrony can positively affect cooperation, even if the cooperative task does not consist of any synchronous action. This mechanism is established early in development at the stage in which young children are learning how to cooperate. Perhaps during music playing, in addition to aligning players, synchrony also prompts players to communicate their intentions better, leading to enhanced performance.

For five-month-old infants, melodies are social

Samuel A. Mehr[1,2], Lee Ann Song[2], and Elizabeth S. Spelke[2]
[1]Graduate School of Education, Harvard University, Cambridge, MA, USA
[2]Department of Psychology, Harvard University, Cambridge, MA, USA

For 1 to 2 weeks, 5-month-old infants listened at home to one of two novel songs with identical lyrics and rhythms, but different melodies; the song was sung by a parent, emanated from a toy, or was sung live by a friendly but unfamiliar adult first in person and subsequently via interactive video. We then tested the infants' selective attention to two novel individuals after one sang the familiar song and the other sang the unfamiliar song. Infants who had experienced a parent singing looked longer at the new person who had sung the familiar melody than at the new person who had sung the unfamiliar melody, and the amount of song exposure at home predicted the size of that preference. Neither effect was observed, however, among infants who had heard the song emanating from a toy or being sung by a socially unrelated person, despite these infants' remarkable memory for the familiar melody, tested an average of more than 8 months later. These findings suggest that melodies produced live and experienced at home by known social partners carry social meaning for infants.

Anytime, Anyplace, Anywhere: The Holy Trinity of Music Streaming?

Geoff Luck[1]
[1]Finnish Centre for Interdisciplinary Music Research, University of Jyväskylä, Jyväskylä, Finland

Digital streaming represents the most radical development in the way we experience music since the invention of automatic playback technologies two centuries ago. From zero ownership and on-demand access to virtually all music ever recorded via a disconnected transaction, the majority of currently available streaming services challenge previous conceptions of how music is defined, experienced and consumed. This paper focuses on the first two of these characteristics, zero ownership and on-demand access, and explores each from a range of psychological perspectives. Part theoretical, part data-driven, this paper combines existing insights with findings from new interview and survey data collected from some 3000 participants across four continents. On-demand access to music we don't own is shown to impact in both positive and negative ways on the three parties around which streaming services revolve: artists, listeners and service providers. Implications and recommendations for the continued co-existence of all three parties in light of the findings are considered.

Is complex coordination more important for maintaining social bonds than synchronization?

Jacques Launay[1], Robin I.M. Dunbar[2]
[1]Department of Life Sciences, Brunel University London, London, UK

While numerous studies suggest that synchronization leads to social bonding it is largely acknowledged that most forms of music and dance do not include exact matching of movements, but instead more complex forms of contingent activity (described as 'coordinated' movement here). Although synchronization can be an effective method of creating bonds between strangers, more complex forms of coordination may be more effective in maintaining existing connections. Coordination involves similar aspects to synchrony (e.g. shared goals, focus of attention, success) but also requires more complex predictions about the movements of other people, and often involves following established rules (e.g. in dance) demonstrating that co-performers share some cultural background. In this experiment participants (n=66) were shown a number of selected YouTube videos in which people either synchronized movements (moving the same time) or coordinated movements (contingent movements without synchronization). Videos involved 'good' performance or 'bad' performance (2x2 design: 'good' vs. 'bad' performance, synchrony vs. coordination). Participants were asked to rate the videos on a series of visual and Likert scales which measured perceived social bonds between the people viewed. Results show that participants only discriminated between videos involving coordination, and perceived closer social bonds between people performing good coordination compared with bad coordination. However, social bonds were not rated differently in the good and bad synchronization videos. Bad coordination was rated as worse than synchrony, but good coordination was rated better. This means that if a group wished to demonstrate strong social bonds they should choose the optimal strategy based on how well they know they can perform together – if their performance could be bad then synchrony is best (e.g. marching), while if they know they can perform well coordination is a better strategy (e.g. dance).

Why listen to music right now? – Towards an inventory measuring the functions of music listening under situational influences

Jochen Steffens[1,2], Fabian Greb[1], Wolff Schlotz[1]
[1]Music Department, Max Planck Institute for Empirical Aesthetics, Germany
[2]Audio Communication Group, Technische Universität Berlin, Germany

Although a large range of factors affecting music listening behavior have already been identified, the question of who listens to what kind of music in which situation and why has not yet been adequately addressed. This paper describes first steps towards an instrument measuring the functions of music listening to predict musical choices in different situations. Our major premise was to capture the wide range of potential functions of music listening. Hence, we reanalyzed data from Schäfer et al. (2013) who did a literature review and extracted 129 distinct functions. For each of the three factors obtained in this study, separate factor analyses were calculated to disclose hierarchically underlying sub-factors. Based on these analyses, an inventory of 22 functions of music listening was created. Since the inventory is intended to capture the situational use of music, questions were formulated accordingly ("Why do you listen to music in this situation?"). The inventory was then used in an online study where participants described three self-selected situations in which they usually listen to music (1761 observations within 587 persons). For each situation, participants also reported on the functions of music listening and actual musical choices. Additionally, they filled out questionnaires on sociodemographics and personality traits. An exploratory factor analysis revealed a five-factor structure of the functions of music listening. The analyses further showed that the factors were substantially correlated with outside variables, in particular with situational factors (e.g. mood, attention) and properties of the chosen music (e.g. tempo, complexity). We conclude that the functions of music serve as relevant predictors of musical choices in different situations. Future research efforts will thus focus on improving the item set and testing its predictive power in a naturalistic setting.

The effect of music listening on Theory of Mind

Jonna K. Vuoskoski[1,2], David M. Greenberg[3,4], Tuomas Eerola[2,5], Simon Baron-Cohen[6], & Jason Rentfrow[3]
[1]Faculty of Music, University of Oxford, UK
[2]Department of Music, University of Jyväskylä, Finland
[3]Department of Psychology, University of Cambridge, UK
[4]Department of Clinical Psychology, City University of New York, USA
[5]Department of Music, Durham University, UK
[6]Autism Research Centre, Department of Psychiatry, University of Cambridge, UK

Manipulations that stimulate empathic engagement and mentalizing – such as reading literary fiction (Kidd & Castano, 2013) – can facilitate performance in tests measuring Theory of Mind. The process of music listening can also involve empathic engagement and narrative imagery (Egermann & McAdams, 2013; Vuoskoski & Eerola, 2012, 2013), and thus it is possible that listening to particular kinds of music could also facilitate Theory of Mind.

Kidd and Castano (2013) postulate that the positive effects of literary fiction on Theory of Mind (compared to reading popular fiction) may have been due to the more complex and nuanced nature of literary fiction, which necessitates the use of mind-reading and perspective-taking abilities. In the present study, we aimed to extend this idea to the context of music listening by comparing two contrasting pairs of musical pieces that differed in emotional expressivity (but not in emotional valence). In Study 1, we compared the effects of an Impressionist orchestral piece to the effects of instrumental ambient music. These two pieces differed in emotional complexity/variability, but both had an overall neutral valence. In Study 2, we compared two different renditions of a Brahms Intermezzo; a MIDI version devoid of any timing variations, and an expressive version performed by a concert pianist.

600+ participants took part in the experiment online via the Musical Universe project, and were randomly assigned to listen to one of the four musical selections. Before and after the music listening, participants completed the revised version of the Reading the Mind in the Eyes Test (RMET; Baron-Cohen et al., 2001), an advanced measure of Theory of Mind comprising 36 pictures (divided into two parts). The Empathy Quotient (Baron-Cohen & Wheelwright, 2004) was used to control for individual differences in trait empathy. Data collection is still in progress, and results will be reported at the conference.

Musical Eclecticism in Terms of the Human Development Index

Michael Barone[1], Matthew Woolhouse[2]
[1] *Psychology, Neuroscience, & Behaviour, McMaster University, Canada*
[2] *School of the Arts, and the McMaster Institute for Music and the Mind, McMaster University, Canada*

Recent sociological research suggests a link between economic and social mobility, and musical eclecticism. Our research uses the components of the Human Development Index (HDI) to explore the relationship between music consumption, and human economic development in various countries across the world. Given its scale—the HDI represents the quality of life in a country with a single number—our research uses large-scale meta-analyses of multiple countries rather than individual case studies. We build upon previous research exploring how music consumption is intimately connected to our working lives (Woolhouse & Bansal, 2013). We examine if the 3 component statistics that comprise the HDI correlate with country-level genre-download preferences. A music-consumption database, consisting of over 1.3 billion downloads onto mobile phones from 2007-14, is used to explore the relationship between the HDI and patterns of music listening. The study was made possible with a data-sharing and cooperation agreement between McMaster University and Nokia Music, a digital music streaming service. The general method, employed primarily for its statistical simplicity, is to correlate genre-dispersion patterns with HDI components, including education, life expectancy, and income. Significant correlations are reported between components of the HDI and genre eclecticism: Countries with higher income and education values tend to download more diverse genres, and are more eclectic in their genre preferences. Other measures related to the HDI, such as the Inequality-Adjusted HDI, demonstrate similar significant correlations. Levels of human economic development, as represented in the HDI, appear to influence the degree to which people are willing to explore diverse musical genres. Cultural eclecticism, or “omnivorousness”, is arguably a function of access to consumption-based leisure, a life quality afforded to individuals living in countries higher on the HDI spectrum.

Empathy and moral judgment: the effects of music and lyrics

Naomi Ziv[1], Shani Esther Twana[2]
[1]Psychology, College of Management – Academic Studies, Rishon Le Zion, Israel
[2]Behavioral Studies, Tel-Aviv Yaffo Academic College, Tel-Aviv, Israel

Previous research has shown that music can increase empathy and affect moral judgment. Similarly, empathy has been shown to reduce severity of moral judgements. However, the role of lyrics in the effect of music has not been systematically controlled. The present study aimed to examine the effects of both music and lyrics on empathy and moral judgment. 120 participants (67 females) were divided into 4 groups and exposed to music with prosocial lyrics, music with lyrics in an unfamiliar language, prosocial lyrics only, or no music or lyrics. They then read a story of a student with ADHD who took a test. His professor did not come to inquire whether he needed any help during the test, as required. When the student later complained about it, the professor accused him of lying. In his frustration, the student then posted an embarrassing photo of the professor on the social network, showing him exiting a pub, drunk. Empathy and evaluation of the student's act were measured. A main effect of music was found, showing both higher empathy and lower severity judgment of his act with music. A main effect of lyrics was found with prosocial lyrics leading to higher empathy ratings. An interaction between music and lyrics was found on both empathy and moral judgment. Whereas lyrics had no effect on empathy in the music groups, they increased empathy in the no-music groups. As for moral judgment, lyrics increased severity of moral judgment in the music groups, but had no effect in the no-music groups. Results suggest that music and lyrics have different effects on empathy and moral judgement. These differences hint at the different processes involved in creating empathy and judging transgressions, suggesting that whereas music affects both in a similar manner, prosocial lyrics increase empathy but may accentuate moral norms, and thus not lead to more lenient moral judgment.

The effects of music on consumer mood states and subsequent purchasing behavior in an online shopping environment

Alison Wong[1], Nicola Dibben[1]
[1]Department of Music, University of Sheffield, Sheffield, UK

Store atmospherics such as music and color are one way retailers try to maximize consumers' feelings of pleasure and willingness to purchase. Musical characteristics such as tempo combined with warm color hues have been found to increase 'desirable' consumer behaviors, but it is unknown to what extent these influences are due to changes in consumer mood nor how they might be mediated by factors such as the character of the product. Our study investigated 3 questions: whether music-induced mood influences measures of consumer behavior online (willingness to pay, likelihood to revisit/repurchase, liking for the online store, general perceptions of the online store); whether congruence between music and color positively influences consumer behavior; and whether these influences are mediated by the hedonic or utilitarian character of the product. We conducted an online experiment using a realistic bespoke online store in a 2 (store color) x 3 (positive high arousal music, positive low arousal music, no music) between-subjects factorial design with one within-subject factor (product category). The results from 226 participants showed that congruence between music and color was associated with an increase in willingness to pay and positive perceptions of the online store. There were effects of music on some measures of mood and consumer behavior and no significant difference in willingness to pay according to product characteristics. These results extend previous findings by showing that arousal states induced by music prior to online shopping can affect consumer behaviors that include willingness to pay, and confirm evidence that shopping behavior is enhanced by congruence between visual and auditory information regardless of product characteristics.

"Gonna make a difference" – Effects of prosocial engagement of musicians on music appraisal of recipients

Nicolas Ruth[1]
[1]Media and Business Communication, Julius-Maximilians-University, Wuerzburg, Germany

There is not much research about effects of prosocial music (music with lyrics about e.g. helping). Many popular musicians deliver their own values through their lyrics and show social engagement. But does the knowledge about this engagement effect the listeners? This study wants to investigate the effects of prosocial music and prosocial behavior of musicians on the recipients' appraisal. The General Learning Model was used as the theoretical base. It describes a continuous cycle between the recipient and the environment and suggests that prosocial content of media may affect a person's internal state by altering their cognition, affect or arousal. An experimental online study (2x2 between subjects design, N = 157, 73% female, M = 24 years) was conducted. As independent variables, two songs (prosocial/neutral) by a newcomer rock band were used and two biographic stories (prosocial/neutral) about the band were written, one focusing on their prosocial work and one without mentioning it. Appraisal (11 items on 5-point Likert scales, α = .89), empathy (9 items on 5-point Likert scales, α = .76), and prosocial behavior (registration for nonprofit organizations) were measured as mediating and dependent variables. The results show that the story effects the appraisal, $F(1,153)$ = 10.81, p <.01, η^2 = .07. The group that read the neutral story and listened to the prosocial song (M = 2.77, SD = .44) showed the most positive appraisal. The group that read the prosocial story and listened to the prosocial song approved the music the least (M = 2.46, SD = .43). This combination seems exaggerated and leads to mistrust towards the band, while a neutral story, where recipients can form their own thoughts about the band, leads to the most positive appraisal of the prosocial song. This could be explained because the newcomer band probably is not regarded credible and authentic. When we read about a band we do not know, we might tend to believe their social work could be showmanship.

A Journey to Ecstasy: The lived experiences of electronic dance music festival attendees

Noah Little[1], Birgitta Burger[1]
[1]Department of Music, University of Jyväskylä, Finland

Recently, there has been a growing global phenomenon of electronic dance music festivals. World-renowned Creamfields, and Miami's Ultra Music Festival, are newly experiencing a drastic influx of festival attendance, societal acceptance, and media coverage. Previously, music festival investigations have primarily focused on motivational factors of attendance, drug incidence, and event management techniques. However, contemporary research has determined attendees are obtaining both psychological and social benefits from these music festivals. This study aimed to provide a detailed exploration of the lived experiences of individuals who attended a multi-day electronic dance music festival and was primarily interested in the perceived beneficial changes within the individual, following their festival experience. A semi-structured qualitative interview was used to collect data from 12 individuals who attended the 2015 Electronic Daisy Carnival in Las Vegas. The data was analyzed using thematic analysis. Within the data emerged the following themes: 1) motivation to attend 2) drug use and abstinence 3) festival atmosphere 4) positive social interactions 5) sense of community 6) perceived beneficial changes. We discovered that individuals believed the festival to be a spiritual-like experience that was sacred and personally meaningful to them. As well, the majority of participants reported abstinence (alcohol, illicit drugs) during their festival experience. Interestingly, following the festival, participants reported the adoption of new values including a greater respect and acceptance for others, the desire to improve themselves, and reduced anxiety. These findings add to the existing body of music festival literature, further contextualizing how music festivals are both experienced, and reflected upon by individuals.

A psychological approach to understanding the varied functions that different music formats serve

Steven Caldwell Brown[*1] & Amanda Krause[#2]

[*]School of Philosophy, Psychology and Language Sciences, The University of Edinburgh, Scotland
[#]School of Psychology and Speech Pathology, Curtin University, Australia
[1]steven.brown@ed.ac.uk [2]amanda.krause@curtin.edu.au

ABSTRACT

There are now a multitude of ways in which consumers can consume and enjoy music. While it is known that most people mix and match between different formats ('multi-channelling' in industry terms), critically, little is known about why people choose to listen to one music format over another. What uses and gratifications do particular formats serve? This study is a pioneering exploration of how consumers negotiate between six major formats: *physical, digital files, free digital streaming, paid-for digital streaming, radio,* and *live music*. It does so in accordance with the theory of planned behaviour, adopting a uses and gratifications approach. Employing an online questionnaire containing measures addressing musical engagement style, music piracy attitudes, the uses and gratifications of preferred listening format, genre preference, personality, and self-efficacy, data was obtained from 396 participants residing in Australia, the US and the UK. The format most used by the sample was by far digital download, followed by streaming, physical, and radio. A factor analysis on the most often used format yielded eight particular uses and gratifications of utilising different formats: *usability and intention to use; discovery; functional utility; flexibility; connection; social norms; value for money*; and *playback diversity*. The presented findings will address how people actively use particular music formats to meet particular needs. Further analyses highlight that different ways of listening to music are related to personality type, engagement in music piracy, and unique ways of engaging with music. Additionally, the study will consider how engagement in illegal downloading fits within a wider pattern of eclectic music-listening behaviours. Discussion centres on the functions that particular formats serve and how consumers tend to listen to music via different formats in complementary ways.

I. INTRODUCTION

Traditionally, recorded music has been purchased legally from retailers on a variety of physical formats – namely vinyl, cassette, and CD. In the last decade, digital mediums have grown in popularity and increasingly music is being listened to online, via computers – with ownership becoming replaced with access. With large libraries of music readily available from music subscription services such as *Spotify* and *Deezer*, it is a curiosity that physical formats continue to shift relatively high volumes of units. For example, vinyl has seen an unprecedented rise in popularity in the last few years.

When choosing to listen to recorded music, consumers have many options related to access choices, such as the type of format, device, and selection methods (Krause & North, 2016a,b). These listening options differ with regard to the degree of control a listener has (e.g., Krause, North, & Hewitt, 2014, 2015), and listener control at least partially stems from *choice*. For example, a personal mobile listening device provides a listener with a different experience than the radio. Music streaming can be thought of as a digital radio of sorts, with users acting as their own personal DJs. For consumers who opt to listen to music via streaming services, a further choice is presented to them – to choose the free or paid-for version. The core difference, besides price, is again control – free versions contain advert-driven interruptions and caps on the amount of time users can listen to music. In terms of price, consumers are of course able to access music illegally – for free. Despite the ever-growing list of legal digital services (over 500), they will always be outnumbered by illegitimate alternatives such as peer-to-peer file-sharing, cyberlockers, and browser plug-ins (to name but a few).

Whilst it has long been assumed that so-called 'music pirates' are interested only in illegally sourced music, a significant body of literature highlights that these same individuals also spend *more* on music legally than those who otherwise do not engage in illegal downloading (Huygen et al., 2009; Karaganis & Renkema, 2013; Watson et al. 2015; Zentner, 2006). Summarizing such research, Schwarz (2014) acknowledges that it is now uncontroversial that this sub-culture are greater consumers of culture overall. The underlying principle here is that those engaging in music piracy have a general thirst for content and indeed individuals scoring high on the personality factor of openness been found to favour music piracy (Brown & MacDonald, 2014).

Illegal sources online offer but one way of accessing music, and the same individuals engaging in illegal downloading also 'mix and match' with other, legitimate services (Sinclair & Green, 2016). Nevertheless, little is known of why people select one format over another. With this in mind, it is curious to note the resurgence of interest in vinyl when music has never been cheaper in human history – it can be accessed for free, in numerous ways online.

It is possible that different formats satisfy different needs beyond mere functionality, and this study aims to empirically explore the varied pros and cons of six different music formats: *physical, digital file, free streaming, paid-for streaming, radio,* and *live music formats*.

II. METHOD

A. Participants

A total of 440 individuals (71.00% female, 28.00% male, 4.00% custom) completed the online questionnaire. Data analysis was completed using the responses from the 396 people who resided in Australia (N = 138), the United States (N = 153), and the United Kingdom (N = 105). Participants were aged 16-71 years (M = 23.53, Mdn = 20.00, SD = 8.98); and 20.70% of the sample held a University degree.

Participants were invited to take part in the study in the first quarter of 2016 and were recruited from University participant pools (in Scotland and Australia), online research websites (e.g., socialpsychology.org), and social media appeals. With the exception of the students who received course credit for taking part via the participant pools, individuals received no compensation for their participation.

B. Materials

1) *Music background and engagement items* Participants were asked to report the average number of hours they listen to music and use technology daily and to rate the importance of music and technology in their lives (on a seven-point scale). They completed Krause and North's (2016a) identity measure and wrote an open-response as to their level of music education/experience. Additionally, participants were asked to name their three favourite artists/bands.

2) *Music Engagement Test (MET; Greenberg & Rentfrow, 2015).* This measure that defines one's listening engagement style using five dimensions (namely cognitive, affective, physical, narrative, and social). The measure includes 23 items and uses a seven-point response scale format.

3) *Personality (Langford, 2003).* This measure asks participants to rate themselves using one item for each of the Big 5 personality dimensions (using a seven-point scale).

4) *Preference and use of six formats.* Using a seven-point scale, participants rated the frequency of their use, their enjoyment, and their self-efficacy for each of six music formats (namely physical, digital file, free digital streaming, paid-for streaming, radio, and live music). Additionally, participants indicated the one format they considered to be their favourite as well as the one format they used most often.

5) *Uses and gratifications.* Individuals were asked to respond to a series of 49 items with regard to the format they used most often to listen to music (seven-point scale). These items were amended from previous research addressing the uses and gratifications of music listening, streaming, and illegal downloading (Krause & North, 2016a; Mäntymäki & Islam, 2015; Sang, Lee, Kim, & Woo, 2015).

6) *Attitudes Towards Music Piracy Scale (AMP-12; Brown & MacDonald, 2014).* On a five-point scale, participants were asked to respond to a series of 12 questions that measure attitudes towards music piracy; higher scores are indicative of more favourable attitudes towards piracy, with attitudes known to be a predictor of engagement (Popham, 2011).

7) *Demographic information.* Participants were asked to report their age, gender, country of residence, and whether they had a University qualification.

C. Procedure

Participants completed a questionnaire, hosted online using Qualtrics. After reading the participant information and indicating their consent to participate, individuals completed the questionnaire as a series of webpages. At the end of the questionnaire, participants were thanked and debriefed.

III. RESULTS

A. Participants

As a sample, the participants considered music to be very important in their lives ($M = 6.45$, $SD = .95$) and listened to music for an average of 3.66 hours daily ($SD = 2.87$). Additionally, 33.10% of the sample considered themselves to be active musicians.

B. Music Format Use

As seen in Table 1, the format most used by the sample was by far digital download, followed by streaming, physical, and radio. Whilst 52.50% of the sample indicated the same format as both their favourite and most often used (46.50% indicated two different formats), it is clear that live music, though the preferred format for almost one third of the sample, it is not frequently experienced.

Table 1. Participant Responses Regarding Format Use

	Response Frequency (Percentage)	
Format	**Favourite**	**Most often used**
Physical	43 (10.90)	25 (6.30)
Digital file	115 (29.00)	163 (41.20)
Free digital streaming	49 (12.40)	83 (21.00)
Paid-for digital streaming	56 (14.10)	88 (22.20)
Radio	12 (3.00)	33 (8.30)
Live music	118 (29.80)	2 (0.50)
No answer	3 (0.80)	2 (0.50)

C. Format Uses and Gratifications

An exploratory principal axis factor analysis with promax rotation on the 49 uses and gratifications items resulted in eight dimensions, which accounted for 54.92% of the variance. Table 2 details the dimensions.

The results support findings from existing research on the role of convenience and functionality of format: the findings highlight the importance of being able to listen to music in ways which match the varied methods of listening to music, such as on the move. The emphasis on discovery is intuitive, given that consumers are now in the position of having access to what often appears like almost every song ever recorded via streaming services. The discovery dimension suggests that a meaningful way of finding music is important to listeners, and new features on subscription services aid this process. The findings of the present study suggest that such features will prove popular, for functional reasons.

Table 2. Uses and Gratifications Dimensions

Dimension	Description	Items (Cronbach's α)
Usability and intention to use	Continued use based on convenience	9 items (.885)
Discovery	Aids discovery of music and broadens taste	11 items (.905)
Functional utility	Storage, ability to control selection	7 items (.891)
Flexibility	Portability, playback matches moment-to-moment demands	5 items (.819)
Connection	Allows user to connect emotionally with music	5 items (.723)
Social norms	Others tend to listen to music in this way	3 items (.875)
Value for money	A financially viable method of listening to music	2 items (.734)
Playback diversity	Format features which allow to shuffle, create playlists	2 items (.596)

The functional utility and flexibility dimensions support previous work which distinguished listening device advantages in terms of a 'progressive' type of advantage, distinct from a 'familiarized' advantage category (see Krause & North, 2016a). Additionally, the desire for value for money (see Brown & Knox, 2016) suggests that participants are methodical in choosing between formats, not mere passive consumers; this highlights that participants choose not only which formats to use for music listening, but which music formats not to use.

D. Additional Analyses

Additional analyses will consider the ways in which participants who prefer different formats differ, in terms of their personality, tendency to engage in music piracy, and musical engagement style.

IV. DISCUSSION

The current study demonstrates why people choose one music format over another, clarifying the varying functions that they serve. The findings highlight the need for further research to unpack the psychological aspects that lead individual listeners to seek out particular music formats and devices. A greater understanding of this has implications not only for theories that apply to understanding the role of music in everyday life but also will benefit the music industry. Importantly, it is clear that consumers mix and match between different formats, with physical and digital formats complementing each other in different ways.

REFERENCES

Brown, S.C., & Knox, D. (2016). Why buy an album? The motivations behind recorded music purchases. *Psychomusicology: Music, Mind, and Brain, 26*, 79–86.

Brown, S. C., & MacDonald, R. A. R. (2014). Predictive factors of music piracy: An exploration of personality using the HEXACO PI-R. *Musicae Scientiae, 18*, 53-64.

Greenberg, D. M., & Rentfrow, P. J. (2015). Rules of engagement: The structure of musical engagement and its personality underpinnings. In J. Ginsborg, A. Lamont, M. Phillips, & S. Bramley (Eds.) *Proceedings of the Ninth Triennial Conference of the European Society for the Cognitive Sciences of Music* (p. 408). Manchester, UK.

Huygen, A., Rutten, P., Huveneers, S., Limonard, S., Poort, J., Leenheer, J., Janssen, K., van Eijk, N., & Helberger, N. (2009). *Ups and downs: Economic and cultural effects of files sharing on music, film and games* (Working Paper No. 1350451). Retrieved from Social Science Research Network website: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1350451

Karaganis, J. and Renkema, L. (2013). *Copy Culture in the US and Germany*. USA: The American Assembly.

Krause, A. E., & North, A. C. (2016a). Music listening in everyday life: Devices, selection methods, and digital technology. *Psychology of Music, 44*, 129-147.

Krause, A. E., & North, A. C. (2016b). Playlists and time perspective. *Psychology of Music*, advanced online publication.

Krause, A. E., North, A. C., & Hewitt, L. Y. (2014). Music selection behaviors in everyday listening. *Journal of Broadcasting and Electronic Media, 58,* 306-323.

Krause, A. E., North, A. C., & Hewitt, L. Y. (2015). Music-listening in everyday life: Devices and choice. *Psychology of Music, 43*, 155-170.

Langford, P. H. (2003). A one-minute measure of the Big Five? Evaluating and abridging Shafer's (1999a) Big Five markers. *Personality and Individual Differences, 35*, 1127-1140.

Mäntymäki, M., & Islam, A. K. M. N. (2015). *Gratifications from using freemium music streaming services: Differences between basic and premium users*. Presented at the Thirty-Sixth International Conference on Information Systems (n.p.).

Popham, J. (2011). Factors influencing music piracy. *Criminal Justice Studies*, *24*, 199–209.

Sang, Y., Lee, J.-K., Kim, Y., & Woo, H.-J. (2015). Understanding the intentions behind illegal downloading: A comparative study of American and Korean college students. *Telematics and Informatics, 32*, 333-343.

Schwarz, J.A. (2014). *Online File Sharing: Innovations in Media Consumption*. New York, NY: Routledge.

Sinclair, G., & Green, T. (2016). Download or stream? Steal or buy? Developing a typology of today's music consumer. *Journal of Consumer Behaviour*, *15*, 3–14.

Watson, S.J., Zizzo, D.J., & Fleming, P. (2015). Determinants of Unlawful File Sharing: A Scoping Review. *PLoS ONE*, *10*: e0127921.

Zentner, A. (2006). Measuring the effect of file sharing on music purchases. *Journal of Law and Economics*, *49*, 63–90.

Picking up good vibrations: The effect of collaborative music making on social bonding and enjoyment

Stephen Ash,*[1] Joshua Albrecht,#[2]

* *Music Department, The University of Mary Hardin-Baylor, United States of America*
[1]soash@mail.umhb.edu, [2]jalbrecht@umhb.com

ABSTRACT

Prior work has shown that music making is reliably correlated with increases in oxytocin and consequently the experiences of social bonding. This study more deeply explores the musical aspects that contribute to the relationship between music making and social bonding. Specifically, we compared a collaborative music making condition in which participants sang with a live accompanist to a non-collaborative condition in which participants sang with a recording. The effects of these conditions were measured against enjoyment, stress, and social bonding. In this study, we report no significant effects of collaboration on physiological measures. Moreover, self-report measures, in which participants provided a rating for questions related to enjoyment, stress, and social bonding, only showed significant effects of collaboration on feeling connected to the accompaniment. However, while most of the physiological and self-report measures were not significant, the results for most of them were skewed in the expected direction.

I. INTRODUCTION

Oxytocin levels in blood have been shown to be directly correlated with several measures of wellness. For example, oxytocin has been implicated in the forgetting of unpleasant memories and the solidification of pleasant memories (Savaskan, *et al.*, 2008). Oxytocin levels are high during pregnancy and child birth, having a role in easing physical pain and discomfort and reinforcing a social bond between mother and child (Leng, *et al.*, 2008). Additionally, oxytocin levels are high during sexual intercourse and during physical contact with loved ones as during hugs, (Winslow *et al.*, 1993) and during teenage years, suggesting that these activities are also prime examples of social bonding (Insel, 1997).

Recently, involvement with music has also been shown to reliably raise blood oxytocin levels (Freeman, 2000; Grape, *et al.*, 2003). Specifically, music listening has been shown to increase oxytocin (Kreutz *et al.*, 2004) as has singing (Keeler *et al.*, 2015). These results suggest that music is a social phenomenon and that part of the function of music making and listening is to solidify social bonds between musicians and between members of the same social group. While the physiological effects of music participation have been shown to be robust, it remains unclear which facets of music are producing these effects. Can experiences of social bonding be manipulated by manipulating the musical experience? Is it the case that certain conditions of music making, for example, produce stronger effects of perceived social bonding than others?

In a study examing the effects of singing on well-being, Grape *et al.* (2003) examined physiological and hormonal levels of singers during a singing lesson. They contrasted professional and amateur singers to determine the different physiological effects produced in these populations by taking singing lessons. They found that both populations exhibited increased cortisol levels, consistent with higher stress levels, as a result of the exertion. However, amateur singers had higher levels of oxytocin and epinephrine levels, suggesting that amateurs experience higher levels of pleasure. This is consistent with the notion that professionals are less inclined to perceive singing as something done to increase wellness. In other words, professionals experience more stress and less pleasure while singing, consistent with the idea that there is more pressure to perform.

II. AIMS

In this study, we contrast two types of music making to examine the effect of social bonding in different circumstances. Specifically, we contrast performance with a recording against collaborative performance to determine whether collaboration leads to greater levels of social bonding than performance with a recording. All factors in both conditions of the experiment were held constant, except that the subjects in the collaborative condition practiced and performed with a live accompanist, while in the recorded condition participants used a recording of the accompaniment. Our hypothesis was that those performing in collaboration with the live accompanist would exhibit increased levels of enjoyment, stress, and perceived social bonding, as indicated by several measures of physiological and self-report data to be collected before, during, and after performance.

III. METHODOLOGY

Although most of the research on this subject has focused on objective measures of social bonding such as oxytocin levels, these measures are relatively expensive and time consuming to collect. Moreover, physiological measures do not necessarily reveal the perceptions of the participant. Therefore, it is appropriate to begin an investigation of the effects of different musical situations on perceptions of social bonding with self-report measures. If, as the prior research suggests, music making is directly tied to social bonding, then the nature of social interactions experienced in music making will likely affect perceptions of social bonding. In other words, if singers make music by themselves instead of with a group of other musicians, we may hypothesize that perceptions of social bonding will be lower.

However, there are many variables that might confound the results of testing a performer singing solo rather than with accompaniment. A better research paradigm would maintain accompaniment conditions while varying the direct interaction with other live performers. In this study, we will contrast two different modes of singing. Both conditions will feature a solo singer with piano accompaniment. However, in the first condition, singers will be accompanied by a recording of a piano accompaniment whereas in the second condition, singers will be accompanied by the same performer who made the recording but in a collaborative environment.

A. Participants

31 undergraduate music majors from the University of Mary Hardin-Baylor participated in the experiment. These participants were randomly divided into two groups of 15 and 16 respectively. 18 of the participants were female, and the average age was 20.4 (SD = 3.3). Participants averaged 97.8 (SD = 14.9) on the Goldsmith's Musical Sophistication Index. Participants were also asked about what their primary instrument was. Of the 31 participants, 13 reported voice as their primary instrument. The average years of vocal instruction were 2.6 (SD = 2.3).

B. Stimuli

All participants sang the same song, "The Black Dress," by John Jacob Niles, transposed to D major. The vocal part of the song is relatively simple, featuring simple contours and rhythms, a range of a minor ninth, and written in a relatively straightforward folk style. Although the texture and melody are not terribly complex, there is nevertheless room for interpretation and rubato. For the purpose of simplicity and time, the first two verses (measures 1-34) were excerpted from the song. The second verse ends with a caesura and a strong authentic cadence.

The pianist who did the accompanying was not aware of the hypothesis of the study nor the experimental conditions. After a period of six days, in which she could familiarize herself with the work, she recorded the song on a Yamaha hybrid model NU1. The performance was recorded directly onto a USB flash storage device in wav format.

In the live performance condition, the accompanist performed live with the singers on the Yamaha hybrid NU1. In the recorded condition, the audio recording was played through the NU1 so that the instrument would be the same in both conditions.

One difficulty with this paradigm is that the accompanist might gradually change her accompaniment style as she becomes more familiar and comfortable with performing the piece. To get an idea of how much her interpretation had changed over the course of the 15 collaborative trials in the experiment, the accompanist was recorded a second time after all trials were completed.

In the first recording, the accompanist took nearly 1:52 for the performance. In the second recording, the accompanist took slightly more than 1:42 to perform. The second recording was 94.6% of the length of the first recording, consistent with the notion that the accompanist may have become more fluent in performing the music over the course of the experiment. However, although this difference is substantial, the effect may have been mitigated by the singers' influence. In other words, accompanists are trained to follow singers, who have a greater role in setting the tempo. Nevertheless, the gradual change of familiarity may be an influence in the results.

C. Procedure

Participants were tested one at a time. Participants arrived in a recital hall, where they were directed through the informed consent process, and given a set of instructions. Each experimental condition had its own set of instructions. Participants read silently along while the instructions were read aloud. Both sets of instructions began with a basic overview of the experiment.

Participants were told that they would be expected to compress the experience of practicing and performing a simple art song into a single session. Each participant was allotted 14 minutes to rehearse the song, and they were encouraged to use the limited time to practice wisely. In the collaborative condition, participants were told that they would rehearse with the live accompanist, while in the recorded condition, they were told that they would rehearse with a CD player equipped with a recording of the accompaniment. In both conditions, participants were encouraged to either ask the live accompanist to pause or focus on special problem areas or to use the capabilities of the CD player to do the same.

After the informed consent process was completed and instructions were read, participants were escorted down the hall to a practice room. In the recorded condition, participants were given a CD player with the audio recording of the performance. Participants were reminded that they only had 14 minutes to practice, and that they could use the CD player to focus on problem areas and the practice room's piano to pick out their melody.

In the collaborative condition, participants and the accompanist were escorted to a practice room. Participants were reminded that they only had 14 minutes to practice, and that they could request the accompanist to focus on problem areas or play through their melody for them. The accompanist was asked before the experiment to not instruct or guide the practice in any way, but just to play or work on whatever the participant wanted, including playing through their sung melody on the piano.

After the 14 minute practice session, the investigator knocked on the practice room door to halt practice activities and escorted the participant (and the accompanist in the collaborative condition) back to the recital hall. Once back in the recital hall, participants were told that blood pressure and heart rate would be monitored during the performance. Participants were fitted with a K18 Heart Rate Smart Wristband on their non-dominant hand, which allows for continuous measure of heart rate.

Once the device was tested to ensure that it was accurately gathering heart rate information, participants completed the Goldsmith's Musical Sophistication Index. Heart rate information during the completion of the survey was used as a baseline heart rate for each participant. After completion of the survey, participants were fitted with the MeasuPro Digital Wrist Blood Pressure Monitor with Heart Rate Meter on their dominant hand. This device uses an inflatable cuff to measure blood pressure and heart rate. The post-survey blood pressure result was used as a baseline, and the heart rate result was used to verify the K18 measurements. The blood pressure monitor was then removed, although participants continued to wear the heart rate monitor.

After completion of the GMSI and the biometric data collection, participants climbed the stage and performed the song. In the collaborative condition, the accompanist and the participant were allowed to briefly discuss last-minute details before performing. In the recorded condition, the experimenter coordinated the initiation of the audio playback through the NU1 piano with the participant.

After the performance, participants' blood pressure was immediately measured again using the blood pressure monitor. Following this test, participants were debriefed in a one-on-one interview with the experimenter. Participants were asked whether they enjoyed the experiment, how easy or difficult they found the task, how familiar they were with the song, whether they knew the name of the composer, how well they thought they performed, how much stress they experienced, how much they enjoyed performing, how connected they felt with the music and accompaniment, and how energetic they felt after the performance.

IV. RESULTS

A. Exclusion criteria

One participant in this study experienced equipment malfunction during the performance, and continuous heart rate data was lost. Additionally, this participant missed every single note except three during the performance. Consequently, the data from this participant was discarded. This participant was in the collaborative condition, and so there were 15 remaining participants per condition.

For another participant, initialized blood pressure data was accidentally not collected. The participant performed well, and heart rate data was collected, so the data we collected was retained. Nevertheless, baseline blood pressure data was missing for that participant, as were comparisons between pre- and post-performance blood pressure.

Table 1. A summary of the physiological results of the study. Differences between collaborative and recorded conditions are shown for simple comparisons and for changes from baseline to performance. As can be seen from the table, none of the differences between conditions was significant.

Metric	Comparison type	Performance Condition	Outcome	Colla-borative	Recor-ded	Difference
Blood pressure	Simple comparisons	Pre	Systolic	130.1	134.6	4.5 (p = .48)
			Diastolic	85.7	80.1	5.6 (p = .23)
		Post	Systolic	131.4	134.7	3.3 (p = .55)
			Diastolic	83.3	82.7	-.6 (p = .89)
	Change from pre- to post-performance		Systolic	-9.9	-0.07	9.83 (p =.29)
			Diastolic	3.3	2.6	0.7 (p =.29)
Heart rate	Simple comparisons	Baseline	Average	82.7	80.6	2.1 (p =.69)
			Maximum	97.8	100.2	(p =.66)
			Minimum	71.3	66.1	(p =.32)
		Performance	Average	84.2	83.2	1 (p =.80)
			Maximum	101.5	98	(p =.57)
			Minimum	69.6	73.1	(p =.41)
	Change from baseline to performance		Average	-1.7	-2.5	.8 (p =.90)
			Maximum	-3.7	-2.2	.5 (p =.34)
			Minimum	-1.7	-7	5.3 (p =.20)

B. Physiological measures

The results attained for the physiological measures are shown in Table 1. In short, using a t-test to examine intergroup differences resulted in no significant differences between conditions for any of the physiological measures. This result held for both simple comparisons (e.g. average heart rate during performance) and for comparisons between changes from that participant's baseline (e.g. difference between average heart rate during performance and average baseline heart rate).

Given the complete lack of any significant effects, it is likely that heart rate and blood pressure are poor substitutes for measuring oxytocin increases or the experience of social bonding in music making. It is also possible that other factors may influence heart rate and blood pressure more than the experimental condition. For example, it may be the case that performing does not raise the heart rate and blood pressure of musicians who have developed coping mechanisms for dealing with stage fright, or for those whose personalities lead them to react minimally to emotionally pressurized situations.

C. Quantitative self-report measures

In addition to the above physiological measurements, participants also answered a series of post-performance follow-up questions. Participants were asked to rate their experiences along a number of dimensions using a 5-point Likert scale.

Table 2. A summary of the quantitative self-reported results of the study. Differences between collaborative and recorded conditions are shown for simple comparisons.

Percieved Reaction	Collaborative	Recorded	Difference
Enjoyment	4.3	3.8	0.5 (p = .094)
Stress	1.77	1.82	0.05 (p = .85)
Quality of performance	3.8	3.5	0.3 (p = .45)
Connection with music	3.5	3.2	0.3 (p = .47)
Connection with accompaniment	4	2.9	1.1 (p =.011)
Energy afterward	4.2	3.6	0.6 (p =.084)

The only rating for which there was a significant difference at $p < .05$ was for how connected participants felt to the accompaniment. In this case, the mean for the collaborative condition was 4.0 compared to 2.9 for the recorded condition ($t(28.0) = 2.6$, $p = .011$). These results are shown in Figure 1.

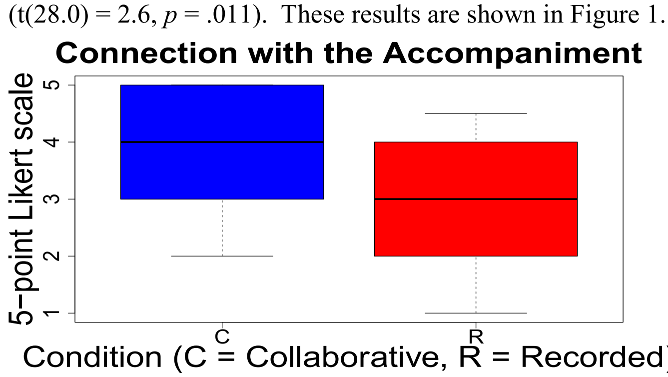


Figure 3. Difference by condition in ratings of "feeling connected with the accompaniment."

Participants in the collaborative condition also reported higher levels of feeling energized (4.2 vs. 3.6) and enjoyment (4.3 vs. 3.8), but those differences were only significant at $p < .1$ ($p = .084$ and $p = .094$, respectively).

Although the data were skewed in the hypothesized directions of collaborative musicians feeling that they performed better (3.8 vs. 3.5), reporting experiencing less stress (1.77 vs. 1.82), and feeling more connected to the music (3.5 vs. 3.2), none of these differences were significant ($p = .45$, $p = .85$, and $p = .47$, respectively). It is possible that with a larger sample size, these differences may be significant at $p < .05$.

D. Qualitative self-report measures

In addition to the quantitative measures on a 5-point scale, participants were also asked for free responses to the same self-report questions. Although quantitative measurements failed to yield significant results for several measures, some of the comments are consistent with the hypothesized effects of collaboration on experiences of social bonding and feeling connected to the music.

For example, when asked how much they enjoyed performing, some participants answers that they enjoyed it very much simply because they like to perform. However, some were less enthusiastic, especially in the recorded condition, including one participant who answered, "I generally don't like performing alone." It is also worth noting that when asked the same question, none of the participants in the collaborative group answered neutrally or negatively. On the contrary, all collaborative condition participants answered that they had enjoyed the experience of performing, though some more emphatically than others.

Collaborative participants' perceptions of how well they performed were varied, but two participants mentioned the accompanist in their answers, though for different reasons. One of them stated, "I had a good connection with the accompanist." The other answered, "I had some disconnection with Mrs. Whitis." In both cases, however, these participants admitted a connection between the quality of their performance and the accompanist. Of the recorded condition's participants, two mentioned that they did not perform as well because they were not able to match the recording when a fermata came in the music.

Although three participants in the recorded accompaniment group reported that they felt connected to the accompaniment, most of the remaining participants did not. Many of the comments related to timing reversals, in which certain moments in the music that typically involve the accompanist following the singer's lead required the singer to follow the recording. For example, one participant said, "The fermatas threw me off badly." Other comments about connecting to the recording spoke directly to the missing human element involved with a recording: "Fine. As good as singing with a recording can be," "Not very, because it's hard to connect with a recording," "Not very. No accompanist to follow," and "Disconnected. It wasn't a real person. When playing with someone, you can see how they breathe. Had to guess where to come in."

Conversely, though not all collaborative group participants admitted to feeling connected with the accompaniment, those who did often mentioned the accompanist, and often did so using the pronouns "she," "her," or "we" instead of "it." Some examples of this were: "She was amazing. She always stayed with me even if I was off with the rhythm or beat. She made it very flowy," "We were connected," "Not hard to follow," "Very connected. Because I am a pianist, so I'm 100% aware of what the accompanist is doing," "It was easy to connect with her," "If I wanted to pick up the tempo, the accompanist followed. We do things musically together," and "She follows me. It's not a background drone. She keeps me going. I don't get too nervous because she's there."

V. DISCUSSION

As mentioned in the results section, there were no significant effects between conditions for either blood pressure or heart rate. Even for the self-report measures, the only significant difference was in feeling connected to the accompaniment. As many of the qualitative explanations reveal, participants are likely not directly thinking about the difference between recorded and live accompanists when answering that question. Because participants are not directly thinking about our research question, it is possible that there are demand characteristics involved in that data.

One possible explanation for the lack of significant difference in heart rate variance under the two conditions could lie with unique effects that music making has on physiological states. Recent research (Vickhoff *et al.*, 2013) has shown that singers' heart rates may sync to musical structures in the music they are singing. Therefore, it is possible that singers' heart rates matched the tempo of the music more than matching their own psychological state.

Another factor that may have confounded our hypothesized results could have been differences in performing strategies between different types of musicians. To investigate this possibility, a k-means cluster analysis on the heart rate and blood pressure results was performed. For these k-means cluster analyses with $k = 2$, solutions showed two distinct clusters of participants for both heart rate changes and blood pressure changes. One cluster of participants experienced a change in average heart rate from baseline to performance of -10 bpm (22 people), and the other a change in average heart rate of +19.5 bpm (8 people). We executed the same clustering analysis for blood pressure, and found that one group had an average change in systolic blood pressure of +12.15 (13 people), and the other group had an average change in systolic blood pressure of -12.38 (16 people). Diastolic blood pressure change also resulted in two groups: +9.5 (14 people) and -10.1 (15 people) (1 participant dropped because of equipment malfunction).

One possible way to explain these results could be that different personality types use different preparation and performance strategies. It may be the case that the participants belonging to clusters exhibiting lower blood pressure and heart rate during performance may not be as influenced by stage fright, while the other group is more influenced by the pressure of performing in front of others, regardless of whether done with collaboration or with a recording. For example, when asked how energetic they felt after the performance, one participant replied, "I had lots of energy after. It's a personality thing."

Intense mental focus and attention are linked with decreased heart rate. It may be the case, then, that participants in the first cluster engaged in strategies that increase mental focus and decrease stress, whereas the second cluster of participants experienced an increase in heart rate and blood pressure due to different strategies that did not as fully block the excitement of the moment.

Furthermore, these different strategies may be linked to differences in personality types. Future research approaches to this question could explore links between physiological reactions and personality, as measured (for example) on the OCEAN Big-5 personality inventory (John & Srivastava, 1999).

It may also be the case that different performance strategies might be tied to levels of musical sophistication or years of vocal training. In a post-hoc analysis, we examined each of these hypotheses by constructing a linear regression model to try to account for differences in average heart rate and blood pressure based on GMSI and years of vocal training.

For our dataset, neither GMSI nor years of vocal training were significantly predictive of blood pressure changes. However, number of years of vocal training was slightly predictive of blood pressure changes (p = .067). This variable was able to account for 8.9% of the variance in average heart rate changes from baseline to performance. Specifically, every year of vocal training increased heart rate during performance by 2.6 bpm.

One possible explanation for this result could be that people who have received training in vocal performance associate performing vocally with nervousness, perhaps because of prior experiences with vocal performance. Another possibility is that participants with vocal training construct their personal identity more around vocal performance. For these participants, any vocal performance may be more important to them because they consider themselves singers. If this is the case, then these singers might experience more stress during performance resulting in higher heart rates.

There may be other reasons why we failed to find any significant effects in the data. In the first case, our sample sizes were relatively small (15 in each condition). Although the data was skewed in the predicted direction in each case, our experiment was underpowered and failed to result in significant differences. A replication with bigger sample sizes may result in significant differences.

Another explanation could simply be that there is no actual effect of collaborative music making on physiological changes. Of course, the primary research question involves levels of oxytocin, which was impossible to include in the current study for logistical reasons. It is therefore entirely possible that musical collaboration might show an effect on oxytocin but not on heart rate or blood pressure.

Further studies could manipulate the tempo of the music to examine the effect on heart rate. Again, conducting the study with bigger groups would be helpful in finding significance for a small effect size. As mentioned above, another fruitful area to explore could be connections with personality characteristics or a more controlled examination of musical sophistication or prior experiences. Additionally, directly examining singers' coping strategies with stage fright could also be a useful future course of study.

VI. CONCLUSION

In this study, results were consistent with the notion that those musicians who perform with another person accompanying them experience a greater feeling of connection to the accompaniment than those who perform alone with a recorded accompaniment. However, no other significant effects were found. Nevertheless, most of the results were skewed in the hypothesized directions. It is possible that a redesigned replication of this study involving bigger samples and a more controlled examination of musical sophistication or performance strategies may result in significant effects of collaboration.

The results of this study also suggest future directions to take this work. For example, it would be worthwhile to further investigate the role of differing personalities and differing levels of experience in music performance on the social bonding, stress, and enjoyment experienced in performance. Further research investigating the effects of musical tempo on participants' heart rates would also be worthwhile.

ACKNOWLEDGMENT

We would like to thank and acknowledge Jessye Whitis for her tireless work in accompanying our participants, making the two recordings, and working around difficult schedules to get this experiment finished. We could not have done this without her, and are deeply grateful for her role in this work.

REFERENCES

Bailey, B.A., & Davidson, J.W. (2002). Adaptive characteristics of group singing: Perceptions from members of a choir for homeless men. *Musicae Scientiae 6*(2), 221-256.

Bailey, B.A., & Davidson, J.W. (2005). Effects of group singing and performance for marginalized and middle-class singers. *Psychology of Music, 33*, 269-303.

Bakker, A.B. (2005). Flow among music teachers and their students: The crossover of peak experiences. *Journal of Vocational Behavior, 66*, 26-44.

Insel, T.R. (1997). A neurological basis of social attachment. *American Journal of Psychiatry, 154*(6), 726-35.

John, O. P., & Srivastava, S. (1999). The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (Vol. 2, pp. 102–138). New York: Guilford Press.

Keeler, J.R., Roth, E.A., Neuser, B.L., Spitsbergen, J.M., Waters, D.J.M., & Vianney, J.M. (2015). The neurochemistry and social flow of singing: bonding and oxytocin. *Frontiers in Human Neuroscience,* 9:518. doi:10.3389/fnhum.2015.00518.

King, E.C. (2006). The roles of student musicians in quartet rehearsals. *Psychology of Music, 34*, 262-282.

Kreutz, G., Bongard, S., Rohrmann, S., Hodapp, V., & Grebe, D. (2004). Effects of choir singing or listening on secretory immunoglobulin A, cortisol, and emotional state. *Journal of Behavioral Medicine, 27*, 623-635.

Leng, G., Meddle, S.L., & Douglas, A.J. (2008). Oxytocin and the maternal brain. *Current Opinion in Pharmacology, 8*(6), 731-734.

Sawyer, R.K. (2006). Group creativity: Musical performance and collaboration. *Psychology of Music, 34*, 148-165.

Savaskan, E., Ehrhardt, R., Schulz, A., Walter, M., & Schächinger, H. (2008). Post-learning intranasal oxytocin modulates human memory for facial identity. *Psychoneuroendocrinology, 33*(3), 368-374.

Roederer, J. (1982). Physical and neuropsychological foundations of music. In M. Clynes (Ed.), *Music, Mind and Brain*. New York and London: Plenum Press.

Roederer, J. (1984). The search for a survival value of music. *Music Perception: An Interdisciplinary Journal, 1*(3), 350-356.

Vickhoff B, Malmgren H, Åström R., Nyberg G, Ekström S.R., Engwall M., Snygg J., Nilsson M. & Jörnsten R. (2013). Music structure determines heart rate variability of singers. *Frontiers of Psychology* 4:334. doi: 10.3389/fpsyg.2013.00334.

Winslow, J.T., Shapiro, L., Carter, C.S., & Insel, T.R. (1993). Oxytocin and complex social behavior: Species comparisons. *Psychopharmacological Bulletin, 29*(3), 409-411.

The effect of music on theory of mind

Adiel Mallik[1], Tabea-Haas Heger[2], Yueyang Li[2] and Daniel J. Levitin[1,2]
[1]Integrated Program in Neuroscience, McGill University, Montreal, Quebec, Canada
[2]Department of Psychology, McGill University, Montreal, Quebec, Canada

Reading literary fiction and music listening have previously been shown to increase empathy. Recently, we showed that when strangers play Rock Band together they have a reduction in social stress and an increase in empathy compared to strangers playing Rock Band individually. Musical videogames may positively affect the mental components of empathy known as affective and cognitive theory of mind (ToM), which could have led to the reduced social stress and increased empathy seen in the previous work. Affective ToM is the ability to detect and understand another person's emotions, while cognitive ToM is the inference and representations of another person's beliefs and intentions. To determine if playing Rock Band increases ToM, healthy subjects (n=44) were recruited and randomly assigned to the experimental group, playing Rock Band (n=22), or to the control group, reading non-fiction (n=22). ToM performance was assessed using the Yoni task and the Positive and Negative Affect Scale (PANAS) was used to assess mood. Yoni is a young boy in a cartoon of various situations. In 1st order Yoni trials; the subject selects the correct object Yoni is referring to, according to the description and Yoni's facial cues. In 2nd order trials the objects are face images and selecting the correct response requires an understanding of the interaction between these images and Yoni's mental state. We found significant ($p < 0.05$) increases in 2nd-order affective trials of the Yoni task, as well as the total affective trial scores and the sum of all scores (1st and 2nd-order affective and cognitive trials) in the experimental group compared to the control group. The PANAS showed no significant changes. This suggests that playing Rock Band improves *affective* ToM but has no discernable effect on *cognitive* ToM. Testing is being done on other experimental groups, to see whether watching the music video, listening to the song, or emotionally neutral video game arousal (Tetris) also affects ToM.

Effects of alcohol-related content in music videos on implicit alcohol identity

Amanda E. Krause[1], Kim M. Caudwell[1], and David A. Keatley[2]
[1] School of Psychology and Speech Pathology, Curtin University, Perth, WA, Australia
[2] School of Psychology, University of Lincoln, Lincoln, United Kingdom

There is increasing research on the prevalence of alcohol-related content in advertising and various media. However, relatively few studies have examined the relationship between alcohol-related content in music videos and alcohol-related cognitions and consumption in non-adolescent samples. The present study investigated the effect of music videos that include alcohol content (i.e., auditory and/or visual) on a measure of implicit alcohol identity (as measured by an implicit association test). Participants completed an online alcohol identity implicit association test before being randomly allocated to one of the experimental conditions. Using a 2x2x2 design, participants were required to attend to a music video that comprised alcohol-related content (yes/no), visual content (i.e., music video; yes/no), and auditory content (yes/no). Participants in the control condition (no music) completed an unrelated task. All participants then repeated the alcohol identity implicit associations test, after which they completed measures of musical preference, music engagement, and alcohol consumption behaviors. We hypothesized that alcohol content in both visual and auditory stimuli would lead to a greater IAT change score than in either stimuli alone, and that there would be no change in a control condition. Primary analyses (underway) compare the pre- and post-manipulation implicit association test scores, with changes in these scores potentially reflective of changes in implicit associations following exposure to alcohol-related content. Additional discussion focuses on the results specific to the participants displaying excessive alcohol consumption behavior as well as interpretation based on different music engagement styles and music genre preferences. Findings are discussed in relation to the potential mechanisms underlying implicit alcohol cognitions and behavior, as well as policy implications surrounding the inclusion of alcohol-related content in music videos.

A light in the dark? The everyday music listening experiences of blind adults and adolescents in the UK

Claire Castle[1]
[1]School of Music, University of Leeds, Leeds, UK

Research provides much support for the previously anecdotal link between visual impairment and musicality, evidencing superior auditory processing abilities in the visually impaired (Röder & Rösler, 2003; Bedny et al., 2011), a greater prevalence of Absolute Pitch in blind musicians than sighted musicians (Hamilton et al., 2004; Gaab, et al., 2006), and a heightened musical interest and ability in visually impaired children (Pring & Ockleford, 2005; Matawa & Ockleford, 2009). However, people with visual impairments are underrepresented in literature relating to everyday musical experiences, despite the increasing prevalence of eye diseases in the UK, and music's role in psychological well-being. It is this gap in the literature that this project seeks to address, with a view to meeting four main aims: to expand on current literature relating to everyday musical experiences; to explore the role of music in the lives of the visually impaired; to assess the accessibility of music for this sub-population; and to identify means of making music more accessible to them. To the researcher's knowledge, this project offers the first systematic consideration of the musical experiences of blind adults and adolescents in the UK.

The project will take place in three stages: focus groups, semi-structured interviews, and a questionnaire. The initial findings outlined here relate to data gathered during a series of focus groups with visually impaired individuals. These sessions were used to explore the role of music in the lives of those with a severe visual impairment, offering opportunities for participants to discuss and compare musical experiences, and a means of identifying barriers to accessing these experiences.

This research contributes to current understandings of the role of music in the lives of individuals and wider society and, through its exploration of the experiences of blind listeners, highlights the importance of including minority and vulnerable groups in music psychological literature.

Music as a facilitator of implicit attitude change

Jonna K. Vuoskoski[1,2], Eric Clarke[1] and Paul Elvers[3]
[1]Faculty of Music, University of Oxford, UK
[2]Department of Music, University of Jyväskylä, Finland
[3]Music Department, Max Planck Institute for Empirical Aesthetics, Germany

In a recent study, Vuoskoski, Clarke, and DeNora (submitted) demonstrated that simply listening to a piece of music from a particular culture (India or West Africa) could evoke implicit affiliation towards members of that culture in general. Furthermore, participants with high trait empathy appeared to be more susceptible to the effect. However, it remains unclear what exact mechanism underlies this effect.

In order to investigate two potential explanations for the effect documented by Vuoskoski and colleagues, two experiments were carried out. In Experiment 1, we investigated the possibility that the affiliation-inducing effect of music listening could simply be due to the activation of positive associations related to the target cultures (e.g., "Indian culture has nice music"). To this end, the same experimental design as used by Vuoskoski et al. was employed, except that instead of music listening, participants viewed a slideshow comprising pleasant images of either Indian or West African musicians. Implicit associations towards the target cultures were subsequently measured using the Implicit Association Test (IAT).

In Experiment 2, we explored the possibility that the effect is driven by empathic resonance and embodied entrainment (as originally hypothesized by Vuoskoski et al.). Participants were randomly assigned to one of four experimental conditions designed to elicit varying degrees of empathic and embodied engagement. Participants were presented with either the original, audiovisual version of an African music performance; a video-only version, an audio-only version, or a manipulated audiovisual version where the audio and video tracks were out of sync. The IAT was used to measure the degree of out-group prejudice (i.e., implicit affiliation for European vs. African people). In both experiments, participants rated the pleasantness of the experimental manipulations, and completed a measure of trait empathy.

Differences between participation in typical *karaoke* singing and *hitokara* singing among college students

Junko Matsumoto[1]
[1]Department of Human Sciences, Nagano College of Nursing, Komagane, Nagano, Japan

Previous studies have reported that *karaoke* singing is associated with positive psychological and physiological effects. It has recently been observed that many young people in Japan go to *karaoke* not only with other people but also by themselves; the latter practice is called *hitokara* (solo *karaoke*). In this study, we investigated the motives underlying participation in both typical *karaoke* and *hitokara*, the mood induced by each activity, and the differences between them. College students in Japan completed a questionnaire on their usual participation in *karaoke* and also in *hitokara*. Results suggested that they engaged in typical *karaoke* for a refreshing change of mood or to have a good time with friends, whereas they opted for *hitokara* for singing practice, refreshing change of mood, and to be able to sing without hesitation. Those with at least one *hitokara* experience felt satisfied and refreshed after *hitokara* singing, and their moods resembled those after *karaoke* singing; however, the intensity level of moods was lower after *hitokara* than that after typical *karaoke*. In addition, they indicated that typical *karaoke* singing was an activity to be enjoyed with other people, whereas *hitokara* singing had different purposes. Consequently, students choose between typical *karaoke* and *hitokara* depending on their purposes or the situations.

Does self-expansion induced by choral singing enhance its psychological and social benefits? A new model of social change

Michał Kierzkowski[1], Rafał Lawendowski[2], Tomasz Besta[2]
[1]Faculty of Choral Conducting, Church Music, Arts Education, Eurhythmics and Jazz, The Stanislaw Moniuszko Academy of Music, Gdańsk, Poland
[2]Institute of Psychology, University of Gdańsk, Gdańsk, Poland

Group singing is one of the most popular and most widely enjoyed form of musical activity. It involves four dimensions: doing, being, belonging, and becoming. Choir participation is beneficial for health and wellbeing, as well as spiritual, emotional, personality, cognitive and social functioning. Regular engagement in group singing enhances mood, helps create positive self-image, increases self-esteem, self-agency and self-efficacy, and teaches strategies for common goal attainment. Identifying and connecting with the choir allows choral singers to form coherent and coordinated groups, while high level of trust enables them to build relationships faster than others do. It could be interesting to determine the mechanism integrating the observed changes visible in the choral singers' self-descriptions, the exhibited attitudes (do they result from personality traits or their social experience?), and the consequences of their functioning, especially in terms of communal actions. The aim of the study was to examine if and how experiencing identity fusion and self-expansion affects the strength of the relation to the group and the sense of subjectivity of a person engaged in group action. Thus, group singers were analysed in terms of their sense of agency and efficacy as well as their perception of out-group members. In the study 110 amateur choral singers were tested longitudinally, thrice in 3 months intervals. The participants completed a questionnaire consisting of the following scales: Identity Fusion, Agency and Communion, Self-Expansion, and items measuring self and group efficacy. The results suggest that self-expansion in the result of choir participation may help explain the relations between group adherence and group-efficacy. Fusion with others could be linked to action tendencies (e.g. stronger engagement in pro-social behaviour) because of the feeling of self-growth related to choir participation.

Comparing musical taste: the effect of background music on taste and preference of cookies

Naomi Ziv[1]
[1]Psychology, College of Management – Academice Studies, Rishon Le Zion, Israel

The pleasure we experience when eating is influenced not just by what we eat, but also by surrounding ambiance and context. In addition, evaluating the taste of two products is influenced by presentation order. Previous studies have shown that background music may affect the experience of taste, with two models – cross-modal correspondence and symbolic priming – explaining this effect. Studies examining order effects on food preference have shown both primacy and recency effects, depending on stimulus conditions. The aim of the present study was to examine the effect of pleasant and unpleasant music on the evaluation and preference for two identical cookies. 60 participants tasted two cookies, 30 participants with pleasant tonal background music followed by atonal unpleasant music, and 30 participants in the reverse order. Participants were asked to evaluate each cookie, and to choose which one they preferred. A main effect of music was found, with the cookie tasted with pleasant music rated as better than the cookie tasted with unpleasant music. No main effect of presentation order was found. An interaction between music and presentation order was found: the cookie tasted with pleasant background music was evaluated as better when tasted first than when tasted second. As for preference, 6 participants expressed no preference. Of the others, chi square showed significant differences in choice of cookie, with higher preference for the first cookie in the group tasting it with pleasant background music. This study is the first to examine presentation order effects of music on the evaluation of two identical products. Although the main effect of music may be explained both by the cross-modal correspondence and by the symbolic priming model, they seem more in line with the latter. Moreover, results suggest that comparing two similar products leads to a primacy effect when the first is accompanied by pleasant music.

ISBN 1-876346-65-5 © ICMPC14

"Let's start giving" – Effects of prosocial music and corresponding media coverage on prosocial behavior

Nicolas Ruth[1]
[1]Media and Business Communication, Julius-Maximilians-University, Wuerzburg, Germany

Little is known about potential effects of music with prosocial lyrics (e.g. about helping). Many popular musicians deliver their own values through their lyrics and dedicate their work to charity projects. Ventures like BandAid come with a lot of media coverage that influences our reception of the music. Thus, the purpose of this study is to investigate the effects of prosocial music and corresponding media coverage on the recipients' appraisal and behavior. Following the General Learning Model any stimulus (including music) is posited to have short-term and long-term effects through several learning mechanisms. It describes a continuous cycle between the recipient and the environment and suggests that prosocial content of media may affect a person's internal state by altering their cognition, affect or arousal. An experimental online study (3x1 between-subjects design, N = 144; 71% female; M = 28.68 years) was conducted to examine how people are influenced by prosocial music that was framed by media coverage. A semi-fictional project was invented: A new version of "We Are the World" is to be released to help the victims of the earthquakes in Nepal in 2015. An online news article about the project was manipulated in three ways: positive, neutral and negative coverage. After reading the story participants watched the music video of "We Are the World" and completed a survey about their appraisal, empathetic feelings and behavior. The results show that only the negative coverage had a negative effect on the appraisal. The appraisal and empathic feelings of the recipients in turn had an influence on their prosocial behavior. It is possible that the neutral and positive media coverage do not affect the recipients because they rely on their own positive or neutral attitude towards the music project, but the negative media coverage gives contradictory information about the project (e.g. musicians are only involved because of monetary reasons) and effects the recipients.

The influence of empowering music on performance and risk-behavior in sports: a randomized controlled trial (RCT)

Paul Elvers[1], Jochen Steffens[1]
[1]Music Department, Max Planck Institute for Empirical Aesthetics, Frankfurt, Germany

Background
It is a common belief that music enhances sport performance and promotes self-esteem in athletes. In sports, music has been used to bring people in the right mindset prior to competitions, helping them to reach the peak of their confidence levels. But does this effect stand firm of scientific scrutiny? And what potential factors might explain an enhancing effect?

Aims
The main objective was to test whether motivational music improves sport performance and promotes risky behavior in sport settings. It was hypothesized that the effect of listening to motivational music is mediated by an enhancement of state self-esteem and a decrease in performance anxiety. It was further assumed that self-selected motivational music would have a greater effect on both outcome variables as compared to experimenter-selected motivational music.

Method
The study design is based on a paradigm developed by DeCharms and Davé. It requires participants to throw a ball into a funnel basket in multiple trials from various distances. While the hitting rate from fixed distances serves as a measure for the throwing performance, risk-behavior is assessed in trials were participants chose distances themselves. Participants were randomly assigned to one of the three experimental groups. The first group listened to self-selected music, the second to experimenter-selected music, and the third served as control group with no music. German versions of the State Self-Esteem Scale (SSES) and the Competitive State Anxiety Inventory (CSAI) were administered as pre- and posttest questionnaires. In order to control for individual differences, ball-throwing skills, self-efficacy beliefs, and trait risk aversion (RA-S) were assessed.

Results
At the time of abstract submission the experiment is being conducted in the laboratory of the Max Planck Institute for Empirical Aesthetics. Data collection is expected to be completed by the end of February.

What do I gain from joining music festival crowds? Does self-expansion help to understand relationship between social identity and functions of music?

Rafał Lawendowski[1], Tomasz Besta[1]
[1]Institute of Psychology, University of Gdańsk, Gdańsk, Poland

Participation in music festivals offers opportunities for the participants to engage in identity work – to define, develop, or reflect on their understanding of themselves, and to cultivate new expressions of self-identity and social identity. Music festival attendees tend to be more open to new kinds of experience and explore new ways of perceiving others. They develop understanding of themselves which gives rise to self-acceptance, self-realization, and personal growth. At the same time, music festival participants have specific preferences for music experiences. Thus far very few attempts have been made to develop a theoretical model integrating the causal relation between these mechanisms. In our study we explore the relation between identity fusion, self-expansion and music experiences. We suggest that the feeling of self-expansion in the result of participation in mass gatherings and the fusion of personal and group identity could be useful in explaining different kinds of psychological experience associated with music. 781 attendees of 3 music festivals participated in our study. They completed a questionnaire consisting of the following scales: Identity Fusion, Agency and Communion, Self-Expansion**,** Functions of Music, and Music Experiences. Based on the structural equation modeling we shown that stronger fusion with fellow festival participants and the inclusion of this group into the self was related to the feeling of self-extension, which results in greater awareness of the things around and the feeling of having a broader perspective on reality. Participants who felt greater self-expansion expressed stronger dedication to the goals of the festival community and used music to facilitate self-awareness, social relatedness, and to regulate mood and arousal. It could be interesting to relate the proposed model to the context of collective behaviour and to different kinds of psychological experience associated with participation in many kinds of communities.

Music in the lives of children and young people with special educational needs and/or disabilities

Sarah Mawby[1], Karen Burland[1], Alinka Greasley[1]
[1]School of Music, University of Leeds, Leeds, UK

It is widely acknowledged that music plays an important role in people's everyday lives. We know that music can be central to an individual's identity and can give pleasure and meaning to our everyday experiences. However, the role of music in the lives of individuals with special educational needs and/or disabilities (SEN/D) has not been sufficiently documented. This paper will discuss ethnographic data from an on-going study into music education for pupils with SEN/D. It will focus on three case studies stemming from the initial year's work (n=3, aged 6-16, profound and multiple learning difficulties). Interviews were conducted with pupils' parents and teachers. An adapted interview with one child was also carried out. This was supported by weekly classroom observations across the autumn term of 2015. Data were analysed using Grounded Theory which facilitated a move from purely descriptive data to theory generation. Results demonstrate that the children showed individual preferences for music, even from an early age. For example, from infancy, on child demonstrated a preference for the indie-folk band Fleet Foxes. Children also used music as a means of mood regulation, often with direct assistance from parents and carers. The complexities of attending out-of-school music activities were also described by parents and teachers, demonstrating a need for more appropriate music opportunities outside school. These results form part of a much wider study which explores best practice in music education for pupils with SEN/D. Nevertheless, they suggest that music in the everyday lives of individuals with SEN/D is worth further investigation. Such research could lead to evidence-based precedents for improved music services for people with complex needs. Furthermore, research in this area could help to make future investigation into music preferences and everyday music experiences more inclusive as people with SEN/D may not be the segregated group we perceive them to be.

When music moves the crowd: typology of music festival participants

Tomasz Besta[1], Rafał Lawendowski[1]
[1]Institute of Psychology, University of Gdańsk, Gdańsk, Poland

Music can be used in various ways in everyday life. It serves different functions depending on time, place, company and individual traits of the listener. Music festivals, which continue to rise in popularity, extend the range of possible music uses. Although literature provides general typologies of listeners, there are still under-researched areas of varieties of music festival participants and the psychological and social consequences of being part of a festival community. As a result, questions regarding the psychological and social consequences of being part of a festival community remain largely unanswered.

The aim of the study was to formulate a typology of participants of three well-known music festivals with established reputation, which are organised in Poland. The criteria of choice were size, profile (highbrow vs. popular) and dominant music genre (mainstream, world music, alternative music). The sample comprised 781 festival attendees, who were examined in terms of correspondence between festival type and three variables: personality profile, music taste, and function attributed to music. The respondents were asked to complete a questionnaire consisting of the following scales: Brief Personality Inventory (TIPI), Identity Fusion with the Crowd, Agency and Communion, Self-Expansion, Music Preferences, and Functions of Music. In the analyses distinct types of music festivals participants were distinguished and grouped into several profiles, both psychological and social. Those profiles were correlated with the festival types and the participants' self-descriptions (self-identity description), the strength of the bond with the festival community. The revealed structures of profiles, associated with both music use and group membership, create comprehensive picture of music festival participants.

Author Index

A

B

C

D

E

F

G

H

I

J

K

L

O

P

Q

R

S

X

Y

Z