# Visualization feedback for musical ensemble practice: A case study on phrase articulation and dynamics

Trevor Knight, Nicolas Bouillot, and Jeremy R. Cooperstock

Centre for Interdisciplinary Research in Music Media and Technology,
McGill University, Montreal, Canada

## ABSTRACT

We consider the possible advantages of visualization in supporting musical interpretation. Specifically, we investigate the use of visualizations in making a subjective judgement of a student's performance compared to reference "expert" performance for particular aspects of musical performance—articulation and dynamics. Our assessment criteria for the effectiveness of the feedback are based on the consistency of judgements made by the participants using each modality, that is to say, in determining how well the student musician matches the reference musician, the time taken to evaluate each pair of samples, and subjective opinion of perceived utility of the feedback.

For articulation, differences in the mean scores assigned by the participants to the reference versus the student performance were not statistically significant for each modality. This suggests that while the visualization strategy did not offer any advantage over presentation of the samples by audio playback alone, visualization nevertheless provided sufficient information to make similar ratings. For dynamics, four of our six participants categorized the visualizations as helpful. The means of their ratings for the visualization-only and both-together conditions were not statistically different but were statistically different from the audio-only treatment, indicating a dominance of the visualizations when presented together with audio. Moreover, the ratings of dynamics under the visualization-only condition were significantly more consistent than the other conditions.

**Keywords:** ensemble musical training, music, pedagogical feedback, visualization, user experiment

## 1. INTRODUCTION

Ensemble musical study allows musicians to practice together and refine interaction with the other instrumentalists, as required for adjusting their playing and accordingly refine their collective style. When performing a collective musical piece, musicians are required to control their playing in order to "fit" with the others, as defined by the particular musical style of ensemble, piece, or even conductor. This introduces several musical considerations and preferences that are not fully specified in musical scores,[1,2] including volume balance, note placement, contrast of dynamics among notes, and the way consecutive notes are articulated. As a consequence, an accurate playing of the score may be considered as appropriate or not, depending on the desired style.

The current investigation was initially motivated by the fact that ensemble practice, however, requires ample time, availability of the other performers, and large spaces, and is therefore less available to student musicians. On the other hand, individual study allows the freedom to practice on any schedule, focusing effort on the specific areas of importance to the performer. Targeting ensemble practice with the flexibility of solo study, our Open Orchestra Project[3] provides an "ensemble simulator" for big band jazz and orchestral instrumentalists at the high school and university level. A simple session description is as follows: a student rehearses along with audiovisual recordings of professional ensembles from the perspective of being in the ensemble and the computer records the audio of the performance. After playing, the student can compare the performance to a reference recording of the original musician playing the same part.

For pedagogical purposes, several studies have shown that instructor demonstration and "recorded models"—recordings of the same part played by an expert—are helpful in the improvement of individual musicians.[4,5] In addition, the importance of self-evaluation and instructive feedback has been demonstrated,[6,7] where the ability

trevorknight@gmail.com, nicolas@cim.mcgill.ca, jer@cim.mcgill.ca

of musicians to listen to, critique, and understand their own playing is regarded as a crucial part of quality musicianship.

Within the Open Orchestra project, the results of the aforementioned studies motivate development of tools that facilitate such self-critique, helping students improve their ensemble performance skills by comparing their own playing to that of an expert,that is to say, the recording of the original ensemble musician playing the same part. Here we investigate two of the musical factors regarded as contributing to ensemble style: articulation and dynamics. These factors were selected as relevant to a musician's nuanced interpretation of their own fit in an ensemble. Although a comparison is possible simply by listening to both audio recordings, we consider whether visualizations may improve the ability of students to do so, either allowing them to observe subtle differences in playing more accurately or faster. However, precise definitions differ between musicians. Thus, a useful visualization strategy must support a variety of subjective interpretation approaches, allowing observers to judge performance at least as effectively as is possible by comparing audio recordings only.

This paper describes our initial effort to exploit visualization as an augmentation of the auditory feedback available by successive playback of two recorded audio segments. We present the design of our visualizations as well as the experiment we conducted to evaluate their effectiveness. The predominant question we consider is whether a visualization feedback strategy allows subjects to judge the quality of "student imitation" as well as, or better than through listening alone. In addition, the results of this experiment help inform subsequent choices of visualization techniques in the Open Orchestra project.

## 2. RELATED WORK

Music visualization methods have been considered for a variety of purposes, ranging from generating a musical score for non-traditional instruments[8] to demonstrating global structure of a piece by examining self-similarity[9] to numerous artistic applications. Our focus, however, is on the use of such visualization for pedagogical purposes, helping students compare their play to that of an "expert" and improve their own performance accordingly.

Existing computer-assisted musical practice software has generally focused on providing feedback to students learning their parts in isolation, analyzing the match between the score and the student performance. The most popular approach is to detect errors and provide a corresponding visualization, helping the student in correcting the play. The Interactive Music Tuition System (IMUTUS) and later the Virtual European Music School (VEMUS)[10,11] were designed to guide a student through lessons and exercises and give feedback about the student's accuracy compared to a displayed score. IMUTUS provides three comments for each student performance based on an analysis of the student's performance and a ranking of the relative importance of different errors. VEMUS is more sophisticated, utilizing the ranking and grading of student errors, creating visualizations of the students' playing, and from the latter, allowing interpretation of timing and dynamics. The visualizations are a form of piano roll notation shown under each line of music, with time on the horizontal axis and pitch lines shown in log scale on the vertical axis. Each note is shown as a polygon, which at any given point is vertically aligned around the pitch with height dependent on the log amplitude of the audio signal. Daudin *et al.*[12] instead represent timbre using stacks of four coloured stripes, whose heights correspond to the amplitude of the corresponding harmonics. Such visualizations for student feedback therefore allow some interpretation of articulation and dynamics from the note shapes and duration. In theory, students could compare the visualizations of their own performance against that of an expert example to observe differences. However, determining areas warranting improvement would require an ability to interpret these differences in performance based on some understanding of the displayed shapes and patterns. This issue does not seem to have been addressed in the literature.

The Digital Violin Tutor (DVT)[13] and its successor, the Interactive Digital Violin Tutor (iDVT)[14] are systems for violin teaching aimed at home computer users, also incorporating performance feedback and performance visualization. Similar to our approach, the systems generate their feedback by comparing expert and student audio files rather than symbolic data. Feedback is also provided by piano roll notation but with blue rectangles for the teacher and green and yellow rectangles for the student. The rectangles show timing and pitch, quantized to the nearest note, thus providing feedback only for gross errors of intonation or fingering. Later work on iDVT[15] integrated video processing as well, utilizing reversals of bowing direction to complement the analysis of acoustic features for onset detection.

In addition to pedagogical applications, other music visualization applications provide useful background and inspiration for the methods presented here. For example, Hiraga and Matsuda[16] developed a visualization method similar to a modified bar graph in an attempt to communicate phrasing and expressiveness of the performance. Onset and duration of each note are indicated by horizontal position and extent, while overall expressiveness of each phrase is shown in greyscale. In an experiment, participants were able to tell the difference between two different phrases, but had difficulty distinguishing two performances of the same phrase.

Sadaka *et al.*[17] created a system for realtime visual feedback, using abstract spiral shapes, to help students mimic changes in timing and loudness. In their experiment, such feedback was found to aid imitation of loudness but not timing. Ferguson *et al.*[18] describe another system for real-time visualization of sound quality, complementary to the traditional loop between a student and a master, showing five musical attributes and allowing the musician to adapt in real-time. Gkiokas *et al.*[19] developed a method for realtime visual feedback of timbre for clarinet players. The system examines the harmonic amplitudes of the notes and classifies them into one of four categories: good, squeak, hollow, or unstable. While an interesting approach for allowing students to see and fix their playing in real-time, this focuses on functional control of the instrument sound, targeting instrument learning rather than ensemble practice, in which awareness of the performance of other musicians is required.

Despite the volume of literature in this domain, there does not appear to be a consensus on representations suitable for facilitating an understanding of the subtleties of one's performance, in particular, highlighting what may be "mistakes" relative to an intended reference example. While we drew inspiration from prior work, we were thus motivated to begin from basics, exploring the fundamentals value of a simple visualization strategy that would expose various aspects of musical performance, one at a time.

## 3. VISUALIZATIONS OF ARTICULATION AND DYNAMICS

Our first guideline was that the user should always have a clear model for comparing their own performance. Since clear and easily interpretable feedback is crucial to musician improvement,[20] we considered that multiple visualizations of specific musical attributes would be easier to learn and understand than a single, more complex, multivariate representation. Our initial focus for visualization and testing was on the two musical features of articulation and dynamics presented. While these terms are defined within music, their interpretation is inherently subjective and perceptual, leaving ambiguity in their associated physical features. This is less of an issue for *dynamics*, which refers to loudness of the playing, but more so for *articulation*, the style of voicing notes—affecting the amount of space between the end of one note and the start of the next as well as their emphasis of notes within a phrase.

The visualization of a student's performance requires a model against which comparisons can be made to pin-point areas in need of improvement. Unfortunately, an objective model, such as one generated from symbolic data, e.g., the music score, lacks the subtlety and nuance of interpretation naturally contained in the audio recording. For this reason, an audio recording is preferred as source data for the visualizations, representing the changes of both articulation and dynamics over time. We performed feature extraction of the RMS values of the recordings with Sonic Annotator[21]) and using Processing* to generate the visualizations.

Figure 1 shows an example articulation visualization of student and reference musicians playing the same nine-note phrase with the reference performance in blue and the student's performance in grey underneath. Time is indicated on the horizontal axis and the height of each shape is proportional to amplitude. In the reference performance, the musician articulates the last two pairs of notes more clearly, whereas the student musician slurs them together. Timing differences between the performances can also be seen on the corresponding two pairs, with the student playing late on the first and early on the second.

Figure 2 shows the same phrase with a different visualization, this time mapping amplitude to both height as well as colour: from yellow to red. The data have also been smoothed with a moving average to focus on overall contour rather than the fine-grained detail. In this view, the crescendo-decrescendo contour of the first five notes of the reference phrase (top) contrasts with the flatter student performance.

*`http://processing.org/`

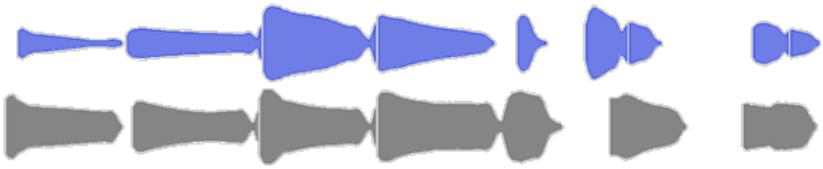


Figure 1. Visualization of articulation with the reference (blue) and student (gray) performance.

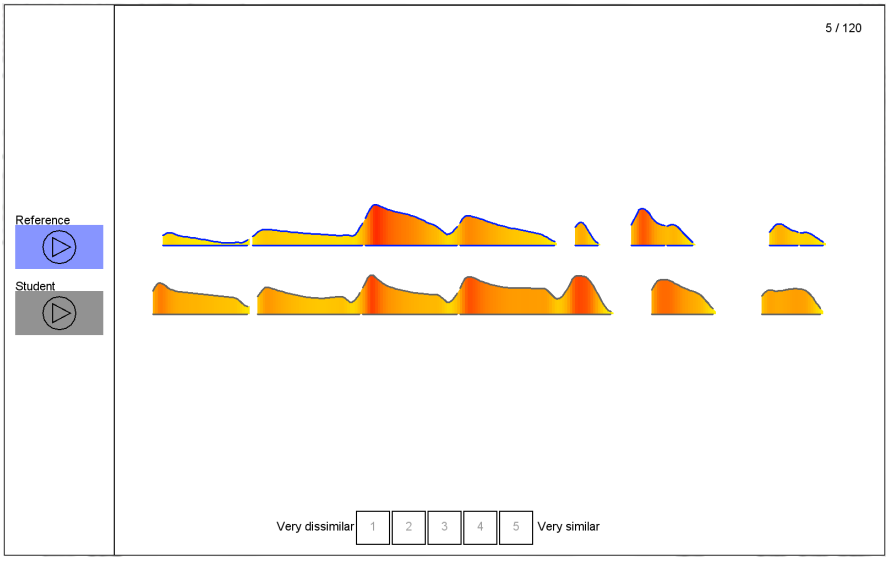



Figure 2. Visualization of dynamics with the reference (top) and student (bottom) performances in the interface used for our experiments.

## 4. EXPERIMENT

To determine the value of different feedback strategies, and specifically, to validate visualization as an effective feedback modality, we conducted an experiment with music students from our target user group. The experimental conditions included audio-only feedback, visualization, and both audio and visualization together. For each of the twenty musical phrases used in the experiment, the participants were asked to rate "How well does the student's [articulation / dynamics] match that of the reference track?" on a five-point Likert scale. The time taken by the participants to rate each phrase was also recorded.

### 4.1 Audio Dataset

The audio dataset was obtained from a previous experiment related to the Open Orchestra project, described by Olmos *et al.*[22] This data consisted of short musical phrases of between 2 and 5 seconds, taken from a longer recording of the first alto saxophone part of a big band jazz piece. The recordings were made of musicians who performed while listening through headphones to a playback of the rest of the ensemble.

The piece was recorded by two saxophonists, one more experienced and well-rehearsed with the piece, while the other was sight-reading the part for the first time that day. The selected phrases spanned a range from nearly identical to significantly different, with wrong notes, poor timing and other errors.

### 4.2 Experimental Design

Six woodwind or brass players with performance and study at the university level participated in the experiment. The participants were seated in front of a computer and asked to asses a "student" performance of a short phrase compared to the "reference", based on a single musical parameter, either articulation or dynamics. Participants evaluated articulation and dynamics on separate days, in balanced order. The experimental interface, as it appeared for the condition of both audio and visualizations, can be seen in Figure 2. For the visualization-only condition, the play button was removed.

For each parameter, the sessions consisted of blocks of the three feedback conditions presented in balanced order across all six permutations. The twenty phrases were presented six times in random order under each condition for a total of 120 phrases per condition. Under the audio-only and audio-visual conditions, the audio for both tracks was played by the participant, sequentially, on-demand, at least once and as many times as desired, before rating.

Participants were given a five-phrase introduction before both the articulation and dynamics portions of the experiment with both audio and visualizations. They were then invited to seek clarification as desired regarding the experiment before proceeding. After completion of each portion of the experiment, participants were asked if they found the visualizations helpful and if so, what aspects and in what ways they were helpful. Additionally, they were asked for any general comments or suggestions for the visualizations with additional follow-up questions as warranted.

### 4.3 Results

Three analyses were performed to examine the effects of the different conditions.

A Friedman analysis of variance test was performed to evaluate the effect of experimental condition on average rating, controlling for question for each dynamics and articulation. For articulation, the responses were not significantly different for the three conditions, but for dynamics, the audio condition was significantly different from visual-only and audio-visual conditions. The visual-only and audio-visual conditions were not significantly different from each other.

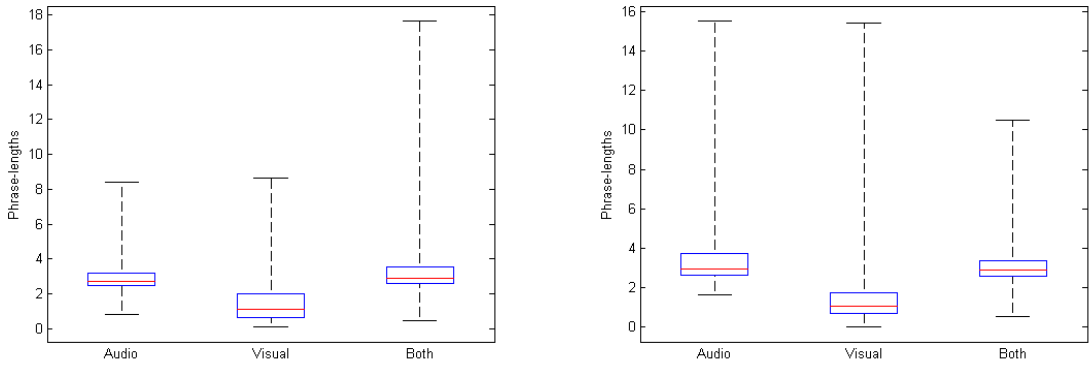
Quite naturally, we were curious as to whether the audio-only ratings were generally more or less "correct" than the other conditions. Given the subjectivity entailed with this question, it is difficult to answer without significantly more evaluation of the segments by musical experts. However, we can gain some insight by considering the consistency of ratings across the six repetitions of each segment under each condition. To do so, we apply an F-test for equality of variances. For articulation, none of the variances of the three treatments were significantly different, but for dynamics, the variance from the visualization-only condition was significantly lower than that of the audio-only and audio-visual conditions (F-test ratio of 1.60, $\alpha < 0.05$). Audio-only and audio-visual variances were not significantly different from each other.

We also analyzed the amount of time participants took to rate a phrase, normalized by the phrase length. In other words, a value of two on the vertical axis means that the participant took twice as long as the length of the phrase to provide a rating. The results indicate that rating times were significantly less for the visualization-only condition for both articulation and dynamics but not significantly different between audio-only and audio-visual conditions.

During the post-test interview, four of the six participants responded that the visualizations of dynamics were helpful, whereas only one of the participants considered the visualization of articulation helpful. Regarding their rating strategy under the combined audio-visual condition, all participants indicated that they used audio across the board. Use of the visualization in the combined condition was correlated with the degree to which the participant considered the visualizations useful.

### 4.4 Discussion

From the fact that all students used audio when available for both articulation and dynamics and not all students found visualizations helpful, visualizations should be provided along with possibility of playing audio recordings. The visualizations may be beneficial, however, when a user wants to review multiple extracts in a short time, where visual feedback allows for faster judgement.




(a) Time taken to rate articulation

(b) Time taken to rate dynamics

Figure 3. Box plot illustrating the amount of time taken to rate a phrase as a ratio of the phrase length. Extents of the box indicate the 25th and 75th percentiles and whiskers extend to extrema.

Table 1. Percentage of participants' ratings within the same phrase and treatment with standard deviations less than 0.753. A greater percentage means lower variation in rating.

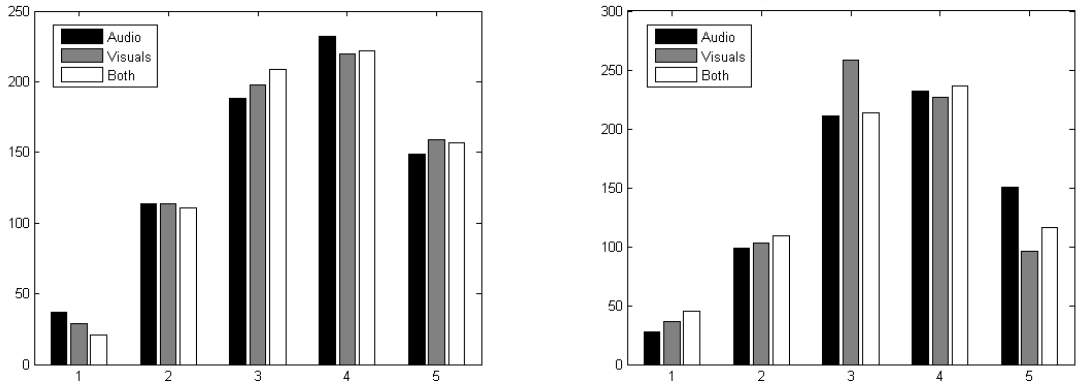
| | Audio-only | Visual-only | Both |
|---|---|---|---|
| Dynamics | 66 | 88 | 73 |
| Articulation | 80 | 81 | 78 |

For the dynamics visualizations, participants mentioned several things they found useful (such as the overall contour, comparison of peaks) even though two users expressed confusion as to what exactly was represented. The participants statements that they found the dynamics visualizations helpful and made use of them during the audio-visual treatment are reinforced by the fact that ratings were significantly different between the audio-only and audio-visual treatments while not being significantly different between audio-visual and visual-only treatments. This even suggests a dominance of the visual modality during the audio-visual treatment and that the visualizations convey the information needed for a comparison. As well, the visualization-only modality allowed faster judgements. As the target users are musicians and all participants used the audio when available, however, a visualization-only feedback strategy may be unadvised.

For articulation, the three groups have equivalent means, suggesting the users obtained a similar rating regardless of modality. While wide variation in rating could account for some of the equivalency of the means, the standard deviations of the ratings are not higher than the articulation experiment. For example, a standard deviation of 0.753 would come from a set of six ratings such as 3, 3, 4, 4, 4, 5. Looking at the standard deviations of participants over the six times they rated the same phrase, 80 percent of the articulation ratings have a standard deviation less than or equal to 0.753. The table 1 shows this percentage broken down by treatment and articulation/dynamics. As well, the distribution of ratings for articulation and dynamics were similar and represented values for the full scale. These distributions are shown in Figure 4.4.

For articulation, even though the means for the three treatments were statistically equal, participants nevertheless reported that they did not find the visualizations helpful. Participants cited many things they were potentially looking or listening for—such as tonguing, fingering, the gestures at the start and end of phrases, timbre, trills, and staccato notes—when judging articulation and four of the six specifically mentioned ambiguity or confusion in what factors they were using. We hypothesize that participants may have found the articulation visualizations unhelpful because they did not represent some of these musical nuances for which they were listening.

For both articulation and dynamics, however, two users had strong feelings about the visualizations even




(a) Rating distribution for articulation (b) Rating distribution for dynamics

Figure 4. The overall distribution of participant ratings along the five-point Likert scale.

if they did not find them helpful, characterizing them as distracting. In fact, the one person who found the articulation visualizations helpful strongly disliked the dynamics visualizations. In light of the strong differences between results for these musical features and the wide range of responses to the visualizations, we must recognize that the success of visualizations is likely to vary greatly based on the content and subjective preferences of each musician. Personalization and options for visualization may therefore be an important aspect of a feedback system built on these visualizations.

The interviews also give some hints at what circumstances the visualizations would be most helpful. For example, one participant suggested that visualizations are most beneficial when the performances are quite similar and therefore useful for identifying just one small point of difference. In a similar vein, several participants commented that when the phrases were too different along several musical parameters, it was difficult to isolate articulation or dynamics either in audio or visual. Two users mentioned visualizations as useful for identifying timing and the duration of rests, therefore note timing may be sufficiently simple to have useful visualizations and would be an interesting direction of future research.

The visualization-only treatment had lower average response times than the other two treatments. In the treatments that included audio, however, the experiment required each participant to at least click the play button for both audio phrases, giving a theoretical minimum speed of two phrase-lengths to finish that rating, having listened to both phrases through to completion. Occasionally, as can be seen in the graph, a participant rated before allowing one of the phrases to finish completely, giving a time-to-rate of under two phrase lengths.

## 5. CONCLUSION

Our small-scale study does not provide any significant results demonstrating that visualization for musician feedback improves the quality of performance assessment. However, the results do indicate that our visualization serves as a useful augmentation to auditory assessment, supporting faster and more consistent ratings. These visualization strategies are implemented within our Open Orchestra system, currently being tested by a large user base. It will be interesting to see how usage, attitudes, and outcomes change as testing progresses.

To increase user satisfaction with the articulation visualizations, it will be necessary to augment the existing visualizations with other data, possibly employing more powerful multivariate representations. In the future, visualizations for a wider range of musical features, such as rhythm and intonation, will be tested and refined with similar experiments, also investigating the benefit of such an approach for expert rating. For more junior musicians, we hope to determine whether visualization might offer advantages over audio alone to help differentiate very similar performances.

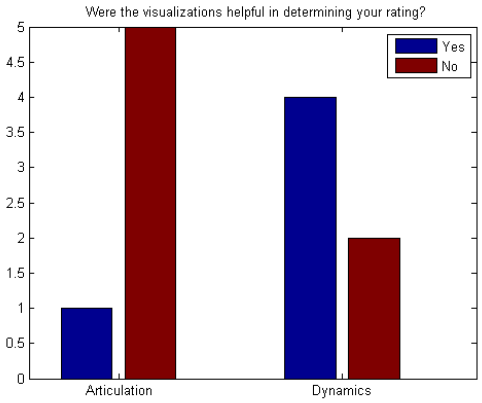



Figure 5. The numbers of participants who found the visualizations helpful.

Furthermore, as the Open Orchestra system aims to replicate an ensemble learning environment, our ongoing efforts will investigate visualizations showing more than just one instrumentalist's part, providing musical context to help the student musicians understand how their part fits with the rest of the ensemble. As seen in the present experiment, generalizations regarding the usability of visualizations may not hold across musical features. The correct or helpful contextual visualizations will therefore vary widely depending on genre, piece, and instrument so intelligently suggested options and configurability will likely be key to usability.

## 6. ACKNOWLEDGEMENTS

The authors would like to thank Ichiro Fujinaga, Adriana Olmos, and Bruno Giordano for their valuable comments and suggestions on experiment design and analysis. The research described here was funded by a student award from the Centre for Interdisciplinary Research in Music Media Technology (CIRMMT) and a Network Enabled Platforms (NEP-2) program research contract from Canada's Advanced Research and Innovation Network (CANARIE).

## REFERENCES

[1] Foote, G. C., *A rhythmic analysis of swing music from the Count Basie library*, Master's thesis, University of Minnesota (July 1979). Master of Arts in Music Education.

[2] Progler, J. A., "Searching for swing: Participatory discrepancies in the jazz rhythm section," *Ethnomusicology* **39**(1), 21–54 (1995).

[3] Olmos, A., Bouillot, N., Knight, T., Mabire, N., Redel, J., and Cooperstock, J. R., "Open Orchestra: A high-fidelity orchestra simulator," *Computer Music Journal* (In press).

[4] Henley, P. T., "Effects of modeling and tempo patterns as practice techniques on the performance of high school instrumentalists," *Journal of Research in Music Education* **49**, 169–180 (2001).

[5] Sang, R. C., "The effects of recorded aural models on the ensemble performance skills of a university concert band," *Bulletin of the Council for Research in Music Education* **91**, 155–159 (1987).

[6] Chaffin, R. and Lemieux, A., [*General perspectives on acheiving musical excellence*], Oxford: Oxford University Press (2004).

[7] Jorgensen, H., [*Strategies for individual practice*], Oxford: Oxford University Press (2004).

[8] Geslin, Y. and Lefevre, A., "Sound and musical representation: The acousmographe software," in [*Proceedings of the International Computer Music Conference (ICMC)*], 285–289 (2004).

[9] Foote, J., "Visualizing music and audio using self-similarity," in [*Proceedings of the 7th ACM International Conference on Multimedia*], 77–80 (1999).

[10] Schoonderwaldt, E., Askenfelt, A., and Hansen, K. F., "Design and implementation of automatic evaluation of recorder performance in IMUTUS," in [*Proceedings of the International Computer Music Conference (ICMC)*], 431–434 (2005).

[11] Fober, D., Letz, S., and Orlarey, Y., "VEMUS - Feedback and groupware technologies for music instrument learning," in [*Proceedings of the 4th Sound and Music Computing Conference (SMC)*], (2007).

[12] Daudin, C., Fober, D., Letz, S., Orlarey, Y., and Chapuis, Y., "Visualisation du jeu instrumental," in [*Actes des Journees d'Informatique Musicale (JIM)*], Grame, ed., 64–72 (2007).

[13] Yin, J., Wang, Y., and Hsu, D., "Digital Violin Tutor: An integrated system for beginning violin learners," in [*Proceedings of the 13th Annual ACM International Conference on Multimedia*], 976–985 (2005).

[14] Wang, Y. and Zhu, J., "Interactive Digital Violin Tutor (iDVT): An edutainment system for violin learners," in [*Proceedings of the International Conference on Advances in Computer Entertainment Technology (ACE)*], 300–301 (2007).

[15] Lu, H., Zhang, B., Wang, Y., and Leow, W. K., "iDVT: An interactive digital violin tutoring system based on audio-visual fusion," in [*Proceedings of the 16th ACM International Conference on Multimedia*], 1005–1006 (2008).

[16] Hiraga, R. and Matsuda, N., "Visualization of music performance as an aid to listener's comprehension," in [*Proceedings of the Working Conference on Advanced Visual Interfaces (AVI)*], 103–106, ACM, New York, NY, USA (2004).

[17] Sadakata, M., Hoppe, D., Brandmeyer, A., Timmers, R., and Desain, P., "Real-time visual feedback for learning to perform short rhythms with expressive variations in timing and loudness," *Journal of New Music Research* **37**(3), 207–220 (2008).

[18] Ferguson, S., Moere, A., and Cabrera, D., "Seeing sound: Real-time sound visualisation in visual feedback loops used for training musicians," in [*Proceedings of the 9th International Conference on Information Visualisation*], 97–102 (2005).

[19] Gkiokas, A., Perifanos, K., and Nikolaidis, S., "Real-time detection and visualization of clarinet bad sounds," in [*Proceedings of the 11th International Conference on Digital Audio Effects (DAFx)*], 59–62 (2008).

[20] Welch, G. F., "A schema theory of how children learn to sing in tune," *Psychology of Music* **13**, 3–18 (Apr. 1985).

[21] Cannam, C., Landone, C., Sandler, M. B., and Bello, J. P., "The Sonic Visualiser: A visualisation platform for semantic descriptors from musical signals," in [*Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR)*], 324–327 (2006).

[22] Olmos, A., Rushka, P., Ko, D., Foote, G., Woszczyk, W., and Cooperstock, J. R., "Where do you want your ears? Comparing performance quality as a function of listening position in a virtual jazz band," in [*Proceedings of the 8th Sound and Music Computing Conference (SMC)*], (2011).