Journal of New Music Research, 2013
Vol. 42, No. 3, 187–204, http://dx.doi.org/10.1080/09298215.2013.788039

Statistical Analysis of Harmony and Melody in Rock Music

David Temperley[1] and Trevor de Clercq[2]

[1]*Eastman School of Music, University of Rochester, USA;* [2]*Ithaca College, USA*

Abstract

We present a corpus of harmonic analyses and melodic transcriptions of rock songs. After explaining the creation and notation of the corpus, we present results of some explorations of the corpus data. We begin by considering the overall distribution of scale-degrees in rock. We then address the issue of key-finding: how the key of a rock song can be identified from harmonic and melodic information. Considering both the distribution of melodic scale-degrees and the distribution of chords (roots), as well as the metrical placement of chords, leads to good key-finding performance. Finally, we discuss how songs within the corpus might be categorized with regard to their pitch organization. Statistical categorization methods point to a clustering of songs that resembles the major/minor distinction in common-practice music, though with some important differences.

1. Introduction

In recent years, corpus methods have assumed an increasingly important role in musical scholarship. Corpus research—that is, the statistical analysis of large bodies of naturally-occurring musical data—provides a basis for the quantitative evaluation of claims about musical structure, and may also reveal hitherto unsuspected patterns and regularities. In some cases, the information needed for a corpus analysis can be extracted in a relatively objective way; examples of this include melodic intervals and rhythmic durations in notated music (von Hippel & Huron, 2000; Patel & Daniele, 2003). In other cases, a certain degree of interpretation and subjectivity is involved. Harmonic structure (chords and keys) is a case of the latter kind; for example, expert musicians may not always agree on whether something is a true 'harmony' or merely an ornamental event. In non-notated music, even the pitches and rhythms of a melody can be open to interpretation. The use of multiple human annotators can be advantageous in these situations, so that the idiosyncrasies of any one individual do not overly affect the results; the level of agreement between annotators can also be measured, giving some indication of the amount of subjectivity involved.

The focus of the current study is on rock music. While rock has been the subject of considerable theoretical discussion (Moore, 2001; Stephenson, 2002; Everett, 2009), there is little consensus on even the most basic issues, such as the normative harmonic progressions of the style, the role of modes and other scale structures, and the distinction between major and minor keys. We believe that statistical analysis can inform the discussion of these issues in useful ways, and can therefore contribute to the development of a more solid theoretical foundation for rock. As with any kind of music, the understanding and interpretation of specific rock songs requires knowledge of the norms and conventions of the style; corpus research can help us to gain a better understanding of these norms.

The manual annotation of harmonic and other musical information also serves another important purpose: to provide training data for automatic music processing systems. Such systems have recently been developed for a variety of practical tasks, such as query-matching (Dannenberg et al., 2007), stylistic classification (Aucouturier & Pachet, 2003), and the labelling of emotional content (van de Laar, 2006). Such models usually employ statistical approaches, and therefore require (or at least could benefit from) statistical knowledge about the relevant musical idiom. For example, an automated harmonic analysis system (whether designed for symbolic or audio input) may require knowledge about the prior probabilities of harmonic patterns and the likelihood of pitches given harmonies. Manually-annotated data provides a way of setting these parameters. Machine-learning methods such as expectation maximization may also be used, but even these methods typically require a hand-annotated 'ground truth' dataset to use as a starting point.

In this paper we present a corpus of harmonic analyses and melodic transcriptions of 200 rock songs. (An earlier version of the harmonic portion of the corpus was presented in de Clercq and Temperley (2011).) The entire corpus is

Correspondence: David Temperley Eastman School of Music, 26 Gibbs St., Rochester, NY 14604, USA. E-mail: dtemperley@esm.rochester.edu

publicly available at www.theory.esm.rochester.edu/rock_corpus. We discuss our procedure for creating the corpus and explain its format. We then present results of some explorations of the corpus data. We begin by considering the overall distribution of scale-degrees in rock, comparing this to the distribution found in common-practice music. We then discuss the issue of key-finding: how the key of a rock song can be identified from harmonic and melodic information. Finally, we discuss how songs within the corpus might be categorized with regard to their pitch organization; an important question here is the validity of the 'major/minor' dichotomy in rock.

Several other corpora of popular music have been compiled by other researchers, although each of these datasets differs from our own in important ways. The Million-Song Dataset (Bertin-Mahieux, Ellis, Whitman, & Lamere, 2011) stands as probably the largest corpus of popular music in existence, but only audio features—not symbolic harmonic and pitch information—are encoded. Datasets more similar to ours have been created by two groups: (1) The Center for Digital Music (CDM) at the University of London (Mauch et al., 2009), which has created corpora of songs by the Beatles, Queen, and Carole King; and (2) The McGill University Billboard Project (Burgoyne, Wild, & Fujinaga, 2011), which has compiled a corpus of songs selected from the Billboard charts spanning 1958 through 1991. Both of these projects provide manually-encoded annotations of rock songs, although the data in these corpora is limited primarily to harmonic and timing information. Like these corpora, our corpus contains harmonic and timing information, but it represents melodic information as well; to our knowledge, it is the first popular music corpus to do so.

2. The corpus

2.1 The Rolling Stone 500 list

The songs in our corpus are taken from *Rolling Stone* magazine's list of the 500 Greatest Songs of All Time (2004). This list was created from a poll of 172 'rock stars and leading authorities' who were asked to select 'songs from the rock'n'roll era'. The top 20 songs from the list are shown in Table 1. As we have discussed elsewhere (de Clercq & Temperley, 2011), 'rock' is an imprecise term, sometimes used in quite a specific sense and sometimes more broadly; the *Rolling Stone* list reflects a fairly broad construal of rock, as it includes a wide range of late twentieth-century popular styles such as late 1950s rock'n'roll, Motown, soul, 1970s punk, and 1990s alternative rock. Whether the songs on the list are actually the *greatest* rock songs is not important for our purposes (and in any case is a matter of opinion). The list is somewhat biased towards the early decades of rock; 206 of the songs are from the 1960s alone. To give our corpus more chronological balance, we included the 20 highest-ranked songs on the list from each decade from the 1950s through the 1990s (the list only contains four songs from 2000 or later, which we did not include). One song, Public Enemy's 'Bring the Noise', was excluded because it was judged to contain no harmony or melody, leaving 99 songs. We then added the 101 highest-ranked songs on the list that were not in this 99-song set, creating a set of 200 songs. It was not possible to balance the larger set chronologically, as the entire 500-song list contains only 22 songs from the 1990s.

2.2 The harmonic analyses

At the most basic level, our harmonic analyses consist of Roman numerals with barlines; Figure 1(a) shows a sample analysis of the chorus of the Ronettes' 'Be My Baby'. (Vertical bars indicate barlines; a bar containing no symbols is assumed to continue the previous chord.) Most songs contain a large amount of repetition, so we devised a notational system that allows repeated patterns and sections to be represented in an efficient manner. A complete harmonic analysis using our notational system is shown in Figure 1(b). The analysis consists of a series of definitions, similar to the 'rewrite rules' used in theoretical syntax, in which a symbol on the left is defined as (i.e. expands to) a series of other symbols on the right. The system is recursive: a symbol that is defined in one expression can be used in the definition of a higher-level symbol. The right-hand side of an expression is some combination of non-terminal symbols (preceded by $), which are defined elsewhere, and terminal symbols, which are actual chords. (An 'R' indicates a segment with no harmony.) For example, the symbol VP (standing for 'verse progression') is defined as a four-measure chord progression; VP then appears in the definition of the higher-level unit Vr (standing for 'verse'), which in turn appears in the definition of S ('song'). The 'top-level' symbol of a song is always S; expanding the definition of S recursively leads to a complete chord progression for the song.

The notation of harmonies follows common conventions. All chords are built on degrees of the major scale unless indicated by a preceding b or #; upper-case Roman numerals indicate major triads and lower-case Roman numerals indicate minor triads. Arabic numbers are used to indicate inversions and extensions (e.g. 6 represents first-inversion and 7 represents an added seventh), and slashes are used to indicate applied chords, e.g. V/ii means V of ii. The symbol '[E]' in Figure 1(b) indicates the key. Key symbols indicate only a tonal centre, without any distinction between major and minor keys (we return to this issue below). Key symbols may also be inserted in the middle of an analysis, indicating a change of key. The time signature may also be indicated (e.g. '[3/4]'); if it is not, 4/4 is assumed. See de Clercq and Temperley (2011) for a more complete description of the harmonic notation system.

Given a harmonic analysis in the format just described, a custom-written computer program expands it into a 'chord list', as shown in Figure 2. The first column indicates the onset time of the chord in relation to the original recording; the second column shows the time in relation to bars. (The downbeat of the first bar is 0.0, the midpoint of that bar is 0.5, the downbeat of the second bar is 1.0, and so on; we call this

Table 1. The top 20 songs on the *Rolling Stone* list of the '500 Greatest Songs of All Time'.

Rank	Title	Artist	Year
1	Like a Rolling Stone	Bob Dylan	1965
2	Satisfaction	The Rolling Stones	1965
3	Imagine	John Lennon	1971
4	What's Going On	Marvin Gaye	1971
5	Respect	Aretha Franklin	1967
6	Good Vibrations	The Beach Boys	1966
7	Johnny B. Goode	Chuck Berry	1958
8	Hey Jude	The Beatles	1968
9	Smells Like Teen Spirit	Nirvana	1991
10	What'd I Say	Ray Charles	1959
11	My Generation	The Who	1965
12	A Change Is Gonna Come	Sam Cooke	1964
13	Yesterday	The Beatles	1965
14	Blowin' in the Wind	Bob Dylan	1963
15	London Calling	The Clash	1980
16	I Want to Hold Your Hand	The Beatles	1963
17	Purple Haze	The Jimi Hendrix Experience	1967
18	Maybellene	Chuck Berry	1955
19	Hound Dog	Elvis Presley	1956
20	Let It Be	The Beatles	1970

A.

```
I | | vi | | IV | | V | |
```

B.

```
In: R | | I | |
VP: I | | ii | V |
Vr: $VP $VP V/vi | | V/ii | | V/V | | V | |
Ch: I | | vi | | IV | | V | |
So: $VP $VP
Fadeout: $Ch
S: [E] $In $Vr $Ch $Vr $Ch $So $Ch R | R | $Fadeout
```

Fig. 1. (a) A harmonic analysis of the chorus of the Ronettes' 'Be My Baby'. (b) DT's harmonic analysis of the Ronettes' 'Be My Baby'.

```
 4.116   2.00     I 0 1 4 4
11.515   6.00    ii 2 2 4 6
13.346   7.00     V 7 5 4 11
15.186   8.00     I 0 1 4 4
18.822  10.00    ii 2 2 4 6
20.682  11.00     V 7 5 4 11
22.490  12.00 V/vi 4 3 4 8
26.207  14.00 V/ii 9 6 4 1
29.866  16.00  V/V 2 2 4 6
33.563  18.00     V 7 5 4 11
```

Fig. 2. Chord list for the first verse of 'Be My Baby', generated from the analysis in Figure 1(b). (The first two bars have no harmony, thus the first chord starts at time 2.0.)

'metrical time'.) The third column is the harmonic symbol exactly as shown in the analysis; the fourth column is the chromatic relative root, that is, the root in relation to the key (I=0, bII=1, II=2, and so on); the fifth column is the diatonic relative root (I=1, bII=2, II=2, etc.); the sixth column shows the key, in integer notation (C=0), and the final column shows the absolute root (C=0).

Each of us (DT and TdC) did harmonic analyses of all 200 songs. We analysed them entirely by ear, not consulting lead sheets or other sources. We resolved certain differences between our analyses, such as barline placement and time signatures. (In most cases, the meter of a rock song is made clear by the drum pattern, given the convention that snare hits occur on the second and fourth beats of the measure; some cases are not so clear-cut, however, especially songs in triple and compound meters.) When we found errors in the harmonic analyses, we corrected them, but we did not resolve differences that reflected genuine disagreements about harmony or key. After correcting errors, the level of agreement between our analyses was 93.3% (meaning that 93.3% of the time, our analyses were in agreement on both the chromatic relative root and the key). Inspection of the differences showed that they arose for a variety of reasons. In some cases, we assigned different roots to a chord, or disagreed as to whether a segment was an independent harmony or contained within another harmony; in other cases, we differed in our judgments of key. (See de Clercq and Temperley (2011) for further discussion and examples.) While it would clearly be desirable to have more than two analyses for each song, the high level of agreement between our analyses suggests that the amount of subjectivity involved is fairly limited.

2.3 The melodic transcriptions

Many interesting questions about rock require information about melody as well as harmony. To investigate the melodic aspects of rock, we first needed to devise a notational system for transcribing rock melodies. Figure 3 shows TdC's melodic transcription for the first verse and chorus of 'Be My Baby' (music notation for the first four bars is shown in Figure 4, Example A). Pitches are represented as scale-degrees in relation to the key: integers indicate degrees of the major scale, while other degrees are indicated with # and b symbols (e.g.

```
[E] [OCT=4] R*4 |
...11.7. | 7671 | ....221. | 7 |
...117.. | 7.67..1. | ..212.1. | 7 |
...3#2.3. | 43.2.3.. | ...3.#23. | 4..3#1.6. |
...2#1.2. | b326. | 7.7.1.12 | ..7.1.22 |
3 | ........3.2.123. | .1...... | ...32122 |
321.2.1......... | .........4..4...3...2... | 4..32..2 |
5432 |
```

Fig. 3. TdC's melodic transcription of the first verse and chorus of the Ronettes' 'Be My Baby'. Figure 4(a) shows music notation for the first four bars.

'#1'). As with our harmonic notation system, vertical bars indicate barlines; each bar is assumed to be evenly divided into units, and a unit containing no pitch event is indicated with a dot. (Only note onset times are indicated, not note offsets.) Each scale-degree is assumed to be the closest representative of that scale-degree to the previous pitch: for example '15' indicates a move from scale-degree 1 down to scale-degree 5 rather than up, since a perfect fourth is smaller than a perfect fifth. (Tritones are assumed to be ascending.) A leap to a pitch an octave above or below the closest representative can be indicated with '^' or '∨', respectively (e.g. '1^5' would be an ascending perfect fifth). Keys and time signatures are indicated in the same way as in the harmonic analyses. [OCT=4] indicates the octave of the first note, following the usual convention of note names (for example, the note name of middle C is C4, thus the octave is 4). From this initial registral designation and the key symbol, the exact pitch of the first note can be determined (given the opening [E] and [OCT=4] symbols, scale-degree 1 is the note E4), and the exact pitch of each subsequent note can be determined inductively (for example, the second note in Figure 3 is the 1 in E major closest to E4, or E4; the third note is D#4). Syllable boundaries are not marked; a sequence of notes will be represented in the same way whether it is under a single syllable (a melisma) or several. 'R∗4' at the beginning indicates four bars of rest before the vocal begins.

For the melodic transcriptions, we divided the 200-song set described above into two halves: DT transcribed 100 of the songs and TdC transcribed the other 100. Six of the songs were judged to have no melody at all, because the vocal line was predominantly spoken rather than sung; for these songs, no melodic data was transcribed. (These include Jimi Hendrix's 'Foxey Lady', the Sex Pistols' 'God Save the Queen', and several rap songs such as Grandmaster Flash and the Furious Five's 'The Message'.) As with the harmonic analyses, we did our melodic transcriptions by ear. We also converted the transcriptions into MIDI files (combining them with MIDI realizations of the harmonic analyses), which allowed us to hear them and correct obvious mistakes. At the website we provide a script for converting an analysis such as that in Figure 3 into a 'note list' (similar to the chord list described earlier), showing the absolute ontime, metrical ontime, pitch, and scale-degree of each note.

For 25 songs in the corpus, both of us transcribed the melody of one section of the song (generally the first verse and chorus), so as to obtain an estimate of the level of agreement between us. In comparing these transcriptions (and in discussing our transcriptions of other songs), we discovered a number of significant differences of opinion. We agreed that we wanted to capture the 'main melody' of the song, but it is not always obvious what that is. Sometimes part of a song (especially the chorus) features an alternation between two vocal parts that seem relatively equal in prominence (usually a lead vocal and a backup group); the chorus of 'Be My Baby' is an example of this (see Figure 4, Example B). In other cases, a song might feature two or three voices singing in close harmony, and again, which of these lines constitutes the main melody may be debatable. In the first bar of Example C, Figure 4, for instance, one vocal part stays on B while the other outlines a descending minor triad. For the 25 songs that we both transcribed, we resolved all large-scale differences of these kinds in order to facilitate comparison; however, we did not attempt to resolve differences in the details of the transcriptions.

Table 2. F scores for DT's and TdC's melodic transcriptions with different values of pitch slack (in semitones) and rhythmic slack (in bars).

	Pitch	
Rhythm	0	1
0	0.893	0.920
1/16	0.906	0.933
1/8	0.953	0.976

After resolving the large-scale differences described above, we compared our 25 overlapping transcriptions in the following way. Given two transcriptions T1 and T2, a note N in T2 is 'matched' in T1 if there is a note in T1 with the same pitch and onset time as N; the proportion of notes in T2 that are matched in T1 indicates the overall level of agreement between the two transcriptions. We used this procedure to calculate the degree to which DT's transcriptions matched TdC's, and vice versa. If we arbitrarily define one set of transcriptions as 'correct' data and the other as 'model' data, these statistics are equivalent to precision and recall. We can combine them in the conventional way into an F score:

$$F = 2/\left((1/\text{precision}) + (1/\text{recall})\right). \tag{1}$$

Roughly speaking, this indicates the proportion of notes on which our transcriptions agree. (If precision and recall are very close, as is the case here, the F score is about the same as the mean of the precision and recall.) The F score for our 23 transcriptions was 0.893. We also experimented with 'slack' factors in both pitch and rhythm. If pitch slack is set at one semitone, that means that a note will be considered to match another note if it is at the same onset time and within one semitone. If rhythmic slack is set at 1/8 of a bar (normally one 8th-note in 4/4 time), that means that a note may be matched by another note of the same absolute pitch within 1/8 of a bar

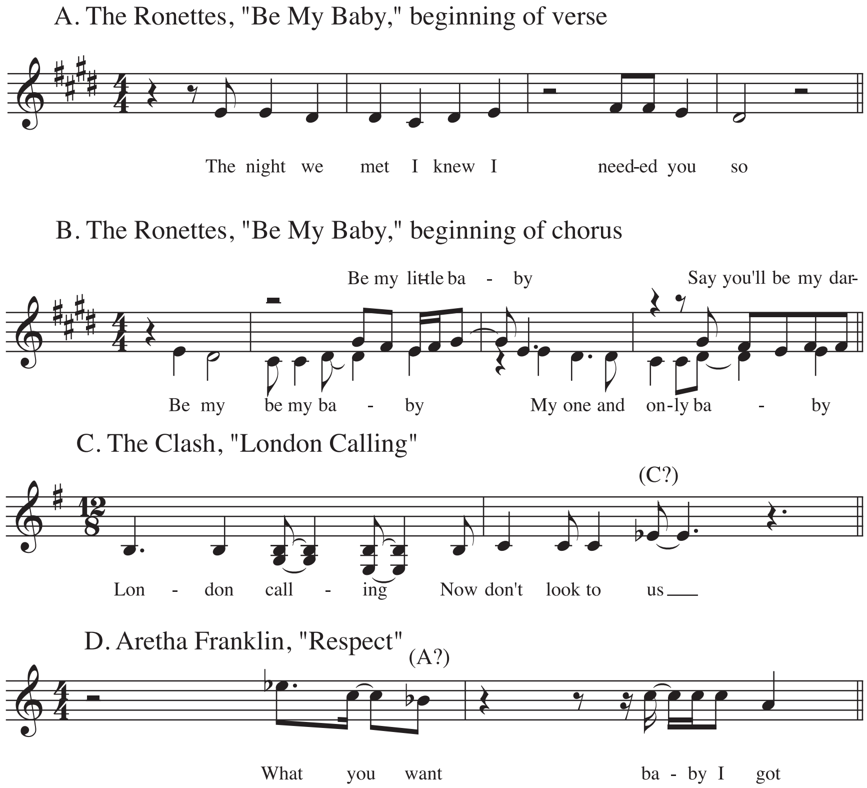

Fig. 4.

of it. *F* scores for various combinations of pitch and rhythm slack values are shown in Table 2.

It can be seen from Table 2 that most of the differences between our analyses can be explained as small discrepancies of pitch and/or rhythm. We inspected these differences to see how and where they arose. In general, we found that TdC tended to take a more ‘literal’ approach to transcription, transcribing exactly what was sung, whereas DT sought to convey the ‘intended’ notes, allowing that these might not correspond exactly to the literal pitches and rhythms. A case in point is seen in the final note in Example C of Figure 4; TdC transcribed this note as Eb4 while DT transcribed it as C4. Literally, this note is approximately Eb4. However, since this melodic phrase is repeated many times throughout the song and generally ends with C4 (scale-degree b6), one might well say that this is the ‘underlying’ pitch here as well, treating the actual pitch as a quasi-spoken inflection added for emphasis. Similar differences arose in rhythm: in some cases, TdC’s transcription reflects a complex rhythm closely following the exact timing of the performance, while DT assumes a simpler underlying rhythm. While we became aware quite early on of this slight difference in approach between us, we decided not to try to resolve it; indeed, we felt that both ‘literal’ and ‘intended’ transcriptions might be useful, depending on the purpose for which they were being used. Another frequent source of differences is ‘blue notes’, i.e. notes that fall ‘between the cracks’ of chromatic scale-degree categories. For example, Aretha Franklin’s ‘Respect’ contains many notes that seem ambiguous between 6 and b7, such as the third note in Example D of Figure 4. We disagreed on several of these notes, though there did not seem to be a particular bias on either of our parts towards one degree or the other.

Despite these differences, the *F* score of 0.893 between our transcriptions showed a high enough level of agreement that we felt confident enough to divide the melodic transcription task between us. Each of our 100 melodic transcriptions was combined with the corresponding harmonic analysis by the same author (DT or TdC) to create a corpus of complete melodic-harmonic encodings for 200 songs. This 200-song collection comprises the dataset used for the statistical analyses described below.

3. Overall distributions of roots and scale-degrees

In the remainder of the article, we present some statistical investigations of our data, and consider what they might tell us about rock more broadly. In de Clercq and Temperley (2011), we presented a smaller version of the harmonic corpus (just the ‘5×20’ portion) and some statistical analyses of that data. Our focus there was on the distribution of harmonies and on patterns of harmonic motion. Here we focus on several further issues, incorporating the melodic data as well.

As a preliminary step, it is worth briefly examining a central issue from our previous study in light of this larger group of

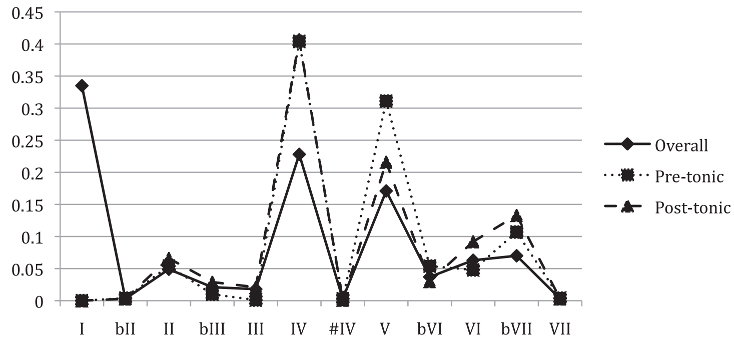

Fig. 5. Root distributions in the rock corpus. For the pre-tonic and post-tonic distributions, I is excluded.

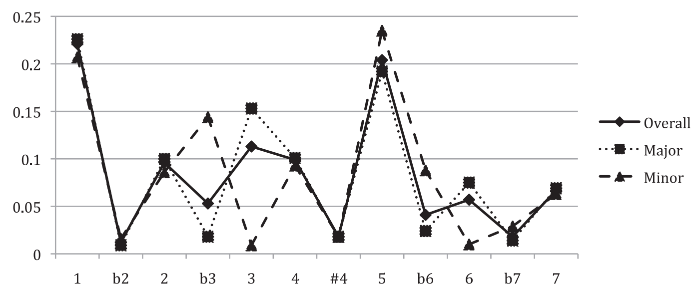

Fig. 6. Scale-degree distributions in a corpus of common-practice excerpts.

songs: the overall distribution of harmonies and scale-degrees. Figure 5 shows the distribution of chromatic relative roots, i.e. roots in relation to the local tonic, for our 200-song dataset. Note here that IV is the most common root after I, followed by V, then bVII, then VI. The graph also shows the proportional frequency of roots immediately before and after I (excluding I); IV is the most common chord in both 'post-tonic' and 'pre-tonic' positions. This data confirms the findings in our previous study of the 5×20 corpus. In general, the progressions of rock reflect a tendency towards temporal symmetry; the strong directional tendencies of classical harmony (e.g. the fact that IV goes to V much more often than the reverse) are not nearly as pronounced in rock.

We now turn to the distribution of scale-degrees, i.e. pitches in relation to the tonic. This issue relates to the question of whether rock has a global 'scale' that favours some scale-degrees over others. Before examining our corpus data, it is interesting to consider similar data from common-practice music (eighteenth- and nineteenth-century Western art music). Figure 6 shows data from a corpus of 46 common-practice excerpts taken from a music theory textbook (Kostka & Payne, 1995; Temperley, 2001). In common-practice music, pieces (or sections of pieces) are typically identified as major or minor; thus Figure 6 shows the data for major and minor excerpts separately. As one might expect, the data from the major-key excerpts strongly reflects the major scale and the data from the minor-key excerpts reflects the harmonic minor scale (containing degrees 1—2—b3—4—5—b6—7); in both profiles, tonic-triad degrees have higher values than other degrees. The figure also shows the data for major and minor excerpts combined. This reflects the union of the major and harmonic minor scales; all twelve scale-degrees are relatively frequent except for b2, #4, and b7. (The b7 degree is a special case; it is part of the descending melodic minor scale, but not nearly as common in minor as 7.) The greater frequency of major degrees over minor degrees is mainly due to the fact that there are more major-key excerpts in the corpus.

Scale-degree distributions can be generated from our rock corpus in two ways. First, we can generate them from the harmonic analyses, by taking each chord to denote a single occurrence of each pitch-class implied by the Roman numeral. For example, I of C major would contain one C, one E, and one G. Unlike the root distribution shown in Figure 5, this distribution incorporates information about chord quality and type (major versus minor, triad versus seventh). However, it does not fully represent the true scale-degree distribution of the corpus. For example, the span of a chord may include many repetitions of a single pitch-class (or doublings in different octaves), or the pitch-class may not occur at all (it may only be implied); the span may also include notes that are not in the chord. Still, this method provides a first approximation. The scale-degree distribution derived from our harmonic analyses is shown in Figure 7, along with the aggregate scale-degree distribution for common-practice music.

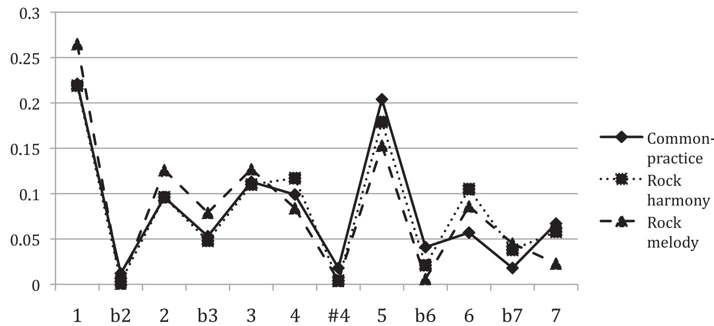

Fig. 7. Scale-degree distributions in common-practice music, rock harmony, and rock melody.

Another way to generate a scale-degree distribution is from the melodic transcriptions; this can be done by simply counting up the occurrences of each scale-degree. This melodic data is shown in Figure 7 as well. It, too, provides only a partial picture of rock's scale-degree distribution, since it ignores the accompaniment, backing vocals, and melodic instrumental lines; but the melodic information does presumably include some non-chord tones, so in this respect it complements the harmonic scale-degree distribution.

It can be seen from Figure 7 that the three distributions—the common-practice distribution, the rock harmony distribution, and the rock melody distribution—are all quite similar. In all three, the most frequent scale-degree is 1, followed by 5. The similarity between the profiles is greatest in the bottom half of the scale. All three distributions can be seen to reflect the primacy of the major scale: 2 > b2, 3 > b3, 6 > b6, and 7 > b7. The one exception, one that we found quite surprising, is that in the melodic distribution—unlike the other two—b7 is more common than 7.[1] We return to this topic below.

Elsewhere, one of us (Temperley, 2001) has proposed that rock is based on a global scale collection containing all 12 scale-degrees except for b2 and #4—the 'supermode'. This concept is supported by the current data. In both the harmonic and melodic rock distributions, the two least frequent scale-degrees are b2 and #4. (This is almost true in the common-practice distribution as well, but b7 is just slightly less common than #4.) This scale could be viewed as the union of the major and natural minor scales; it also corresponds to a set of ten adjacent positions on the circle of fifths. We should note, however, that in the melodic data, b6 is only slightly more common than b2 and #4; a classification of scale-degrees into 'scalar' and 'chromatic' might well place b6 in the latter category.

It might be argued that these aggregate scale-degree distributions tell us little, since most actual rock songs use only a subset of these degrees; that is, they employ some kind of smaller scale collection within the 'supermode' (just as most classical pieces use the major or minor scales rather than the union of them). But what exactly are these smaller scale collections? We return to this issue in Section 5 below.

4. Key-finding

4.1 Background

How do listeners determine the key of a rock song as they hear it? This is a basic and important question that can be asked about any kind of tonal music. The problem of 'key-finding', as it is sometimes called, has received considerable attention in music psychology (for a review, see Temperley, 2007b). Key-finding has practical implications as well, since many music-processing tasks depend on having the correct tonal framework. For example, in identifying one song as a version or 'cover' of another, what is crucial is not their melodic or harmonic similarity in absolute terms—the two songs may well be in different keys—but rather, their similarity in relation to their respective keys (Serrà, Gómez, & Herrera, 2010).

Key-finding models can be classified as to whether they accept audio or symbolic information as input. Clearly, audio-input models more directly reflect the situation of human listening. However, much of the information in an audio signal is presumably of little relevance to key—information such as timbre, percussion parts, small nuances of pitch (e.g. vibrato), and lyrics. A reasonable approach to key-finding would be to extract the information that is crucial for key identification—primarily, categorical pitch information—and then perform key-finding on that reduced representation. Indeed, some audio key-finding models have taken exactly this approach (Cremer, 2004; Pauws, 2004). Symbolic key-finding models—such as the ones we present below—could be seen as proposing solutions to the second part of this process.

Most symbolic key-finding models have employed what could be called a distributional approach. Each key is associated with an ideal distribution of pitch-classes (known as a 'key-profile'), and the key whose distribution most closely

[1]In a large corpus of European folk music, the Essen Folksong Collection (Schaffrath1995; Temperley, 2007), b7 is more common than 7 in minor-key melodies, though 7 is more common than b7 overall.

matches that of the piece is the preferred key; this is the essential idea behind the classic Krumhansl–Schmuckler key-finding model (Krumhansl, 1990) and several later models (Vos & Geenen, 1996; Shmulevich & Yli-Harja, 2001; Temperley, 2007b). The approach of Chew (2002) is similar, though both keys and pieces are represented by points in a spiral spatial representation. For audio key-finding, explicit pitch information is of course not available; many models instead use *chromas*, octave-spaced frequency components in an audio signal corresponding to pitch-classes (Purwins, Blankertz, & Obermayer, 2000; Chai & Vercoe, 2005; Gomez, 2006; Izmirili, 2007). A characteristic chroma profile can be created for each key, and this can be matched to the chroma content of the piece in a manner similar to key-profile models. Some studies have addressed the more difficult problem of deriving both harmonic structure and key simultaneously; such models are able to take into account the interdependence of harmony and key (Raphael & Stoddard, 2004; Lee & Slaney, 2007; Mauch & Dixon, 2010; Rocher, Robine, Hanna, & Oudre, 2010; Papadopoulos & Peeters, 2011).

While most of the above-mentioned studies focus on classical music, several of them explore key estimation in popular music. Most of these employ audio input. Lee and Slaney (2007), Mauch and Dixon (2010), and Rocher et al. (2010) test their models on the CDM Beatles corpus, mentioned earlier; Izmirli (2007) and Papadopoulos and Peeters (2011) use both classical and popular materials. Very few models have addressed the specific problem addressed here, namely, the identification of key in popular music from symbolic information. One study deserving mention is that of Noland and Sandler (2006). This model estimates key from the transitions between chords (labelled as major or minor triads) and is tested on the CDM Beatles corpus; the model identifies keys with 87% accuracy.

In what follows, we explore a variety of methods for identifying key in popular music, using our melodic transcriptions and harmonic analyses as input. We begin by considering several models that employ pitch-class distributions; we then consider ways of estimating key directly from harmonic information. All of our models use the same basic approach to parameter-setting and testing. The set of 200 melodic-harmonic analyses described above (100 by DT, 100 by TdC) was randomly split into two sets of 100; one set was used for training (parameter-setting) and the other set was used for testing.

Most work on key estimation in popular music has identified keys as major or minor, following the common-practice key system. However, we found in creating our corpus that it was often quite problematic to label songs as major or minor (we return to this issue in Section 5). Thus, we simply treat a ‘key’ in rock as a single pitch-class. Another issue that arises here concerns modulation, that is, changes of key within a song. While most rock songs remain in a single key throughout, some songs change key: 31 of the 200 songs in our corpus contain at least one key change. All songs containing modulations were placed in the training set; thus all songs used for testing contain just one key. The identification of modulations within a song is an interesting problem, but we do not address it here.

4.2 Pitch-based key-finding

The most obvious way to apply a pitch-based key-finding strategy given our corpus is simply to use the melodic transcriptions. A key-profile—a normative distribution of scale-degrees—can be generated from the melodic transcriptions in the training set. The resulting profile is very similar to the melodic scale-degree distribution shown in Figure 7 (it is just slightly different since it is based on 100 songs rather than 200). Transposing the profile produces a pitch-class profile for each key. The key of each song in the test set can then be chosen by finding the pitch-class distribution of the song, which we call an input vector (following Krumhansl, 1990), and choosing the key whose profile best matches the input vector. We measure the similarity between a key-profile and an input vector using a probabilistic method proposed by Temperley (2007b). If all keys are equal in prior probability:

$$\mathrm{argmax}_{\mathrm{key}} P(\mathrm{key}|\mathrm{song}) = \mathrm{argmax}_{\mathrm{key}} P(IV|\mathrm{key}) \quad (2)$$

$$= \mathrm{argmax}_{\mathrm{key}} \log P(IV|\mathrm{key})$$

$$= \mathrm{argmax}_{\mathrm{key}} \sum_{\mathrm{pc}} IV(\mathrm{pc}) \log KP_{\mathrm{key}}(\mathrm{pc}), \quad (3)$$

where IV is an input vector, KP is a key-profile, pcs are pitch classes, $IV(\mathrm{pc})$ is the input-vector value for a pc, and $KP(\mathrm{pc})$ is the key-profile value for a pc. The final expression above can also be viewed as the (negative) cross-entropy between the key-profile and the input vector. A similar matching procedure is used in all the models presented below.

We call this Model 1; it is essentially equivalent to the monophonic key-finding model proposed in Temperley (2007b). Table 3 shows the model’s performance on the 100-song test set. We also reasoned that it might help to weight each note according to its duration, as in Krumhansl’s (1990) model. This cannot easily be done in our data set, since only note onsets are indicated, not offsets. In general, however, the offset of a melodic note approximately coincides with the next note onset (it could not be *later* than the next onset, given the limitations of the human voice), unless the time interval between the two onsets is very long; in the latter case, there is often a rest between the two notes (this frequently happens at phrase boundaries, for example). Thus we define the length of a note as its inter-onset interval (the time interval between its onset and the next, in metrical time) or one bar, whichever is less. With notes thus weighted for duration, key-profiles and input vectors can be constructed as they were in Model 1. This is Model 2; as shown in Table 3, it performs somewhat better than Model 1. Since long notes tend to occur at the ends of phrases, the superior performance of Model 2 may in part be due to the fact that it is giving extra weight to notes in phrase-final positions.

Table 3. Key-finding models.

Model	Description	Number of parameters	% correct
1	Melodic scale-degree (SD) distribution	12	76
2	Melodic SD distribution, weighted for duration	12	82
3	Harmonic SD distribution, weighted for duration	12	78
4	Weighted mel. SD dist. + weighted harm. SD dist.	24	86
5	Root distribution	12	77
6	Root dist. weighted for duration	12	86
7	Root dist. distinguishing metrically strong vs. weak	24	89
8	Root dist. weighted for duration, distinguishing strong vs. weak	24	91
9	Root dist. weighted for duration, distinguishing strong vs. weak, + weighted melodic SD dist.	36	97

Models 1 and 2 only use melodic information, and therefore consider only part of the pitch-class content of the piece. Another approach to pitch-based key-finding is to use a harmonic analysis to generate an approximate distribution of pitch-classes—taking each chord to imply one instance of each pitch-class it contains, as was done in Section 3 above (the resulting profile is very similar to the rock harmony distribution in Figure 7). As with our melodic models, we found that weighting each event by its length improved performance; in this case, the length of a note is defined by the length of the chord that contains it, which is explicitly indicated in our harmonic analyses. Key-profiles and input vectors can then be generated in the usual way. The results of this model (Model 3) are shown in Table 3; the model performs slightly less well than the duration-weighted melodic model.

As noted earlier, the melodic and harmonic scale-degree distributions are in a sense complementary. The harmonic distribution includes accompaniment parts and instrumental sections, but excludes non-harmonic notes; the melodic distribution includes such notes but considers only the melody. A logical next step, therefore, is to combine the two distributions, creating a profile with 24 values. This model, Model 4, performs slightly better than Model 2, which considers melodic information alone. We earlier noted some differences between the harmonic and melodic key-profiles, notably the greater frequency of b7 in the melodic profile; recognizing this distinction appears to have a small benefit for key-finding.

4.3 Harmony-based key-finding

Even the best of our pitch-based key-finding models leaves considerable room for improvement. An alternative approach is to estimate key directly from harmonic information. Noland and Sandler (2006) employ this strategy, with good results; their model looks at transitions between chords—the assumption being that important information about key might lie in the temporal ordering of chords. We have found, however, that the strong ordering constraints of classical harmony are much less present in rock: for example, the frequency of the IV chord does not seem to vary greatly depending on its position in relation to the tonic (see Figure 5). Thus we doubted that considering chord transitions would give much boost to performance. Instead, we simply represent harmonies in a 'zeroth-order' fashion: the probability of a chord occurring depends only on the key and not on the previous chord or chords.

```
a.   C . | . G | C . | . G |
b.   G . | . C | G . | . C |
c.   C G | C G | C G | C G |
d.   G C | G C | G C | G C |
e.   C G | . . | C G | . . |
```

Fig. 8. Five chord progressions. Capital letters represent major chords.

Our first harmony-based model simply considers the overall distribution of roots. This is essentially a key-profile model, similar to those presented above, but the key-profile in this case represents the frequency of each root in relation to the tonic—very similar to the 'overall' root distribution shown in Figure 5. The input vector represents the root distribution of a song in an absolute fashion, and the usual cross-entropy method is used to find the best-matching key given the input. This is shown as Model 5 in Table 3. One limitation of Model 5 is that it does not consider the durations of chords. Consider progressions such as those in Figures 8(a) and (b); it seems to us that C is a more likely tonic in the first case, and G in the second, though by Model 5 the two progressions are considered equivalent (since both contain two C chords and two G chords). Thus we tried weighting chords by duration (as we did with the melodic algorithms presented earlier). We do this by treating each half-bar segment as a separate harmonic token (whether it is the beginning of a harmony or not), labelled with the harmony that begins or is in progress

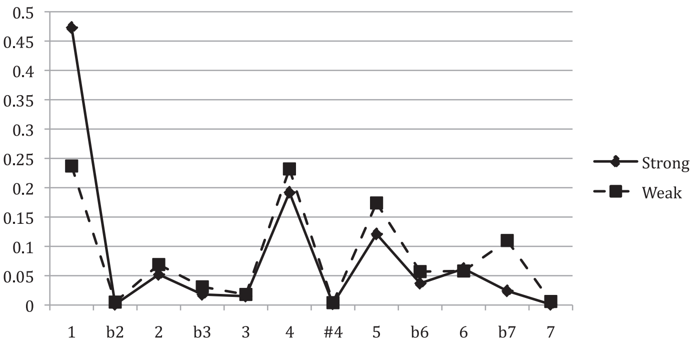

Fig. 9. Key-profiles for metrically strong and weak harmonies.

at the beginning of the segment; the key-profile and input vector represent counts of these tokens. (About 3% of chords in our analyses do not span any half-bar point; these are effectively ignored.) Thus a harmony that extends over several half-bar segments will carry more weight. The resulting model is shown in Table 3 as Model 6.

We suggested that the preference for C over G as tonal centre in Figure 8(a) might be due to the greater duration of C harmonically. Another factor may be at work as well, however. Some authors have suggested that the metrical placement of harmonies plays an important role in key-finding in rock: there is a strong tendency to favour a key whose tonic chord appears at metrically strong positions (Temperley, 2001; Stephenson, 2002). This would have the desired result in Figures 8(a) and (b) (favouring C in the first case and G in the second) but it would also distinguish Figures 8(c) and (d) (which Model 6 would not), favouring C in the first case and G in the second, which we believe is correct. Model 7 is similar to Model 5, in that it counts each chord just once, but it maintains one key-profile for chords beginning at metrically strong positions and another profile for other chords, and also distinguishes these cases in the input vector. We experimented with different ways of defining 'metrically strong'; the most effective method was to define a strong position as the downbeat of an odd-numbered bar (excluding any partial bar or 'upbeat' at the beginning of the song). The profiles for weak and strong harmonies are shown in Figure 9; it can be seen that, indeed, the main difference is the much higher frequency of tonic in the metrically strong profile. (The data in Figure 9 also indirectly confirm our assumption that, in general, odd-numbered bars in rock songs are metrically strong. Many rock songs are composed entirely of four-bar units in which the first and third bars of each unit feel metrically stronger than the second and fourth. There are certainly exceptions, however; a number of songs in our corpus contain irregular phrases, so that the correspondence between odd-numbered bars and strong downbeats breaks down. No doubt a more accurate labelling of metrically strong bars would improve the performance of Model 7, but we will not attempt that here.)

Model 7 yields a slight improvement over previous models, but further improvement is still possible. Model 7 considers only the starting position of chords, not their length; therefore the two progressions in Figures 8(c) and (e) are treated as equivalent, since in both cases, all the C chords are strong and all the G chords are weak. But it seems to us that the greater duration of G in Figure 8(e) gives a certain advantage to the corresponding key. Model 8 combines Models 6 and 7: we count each half-bar separately, but use different distributions for 'strong' half-bars (those starting on odd-numbered downbeats) and 'weak' ones. The result on the test set is 91% correct, our best result so far.

As a final step, we experimented with combining the harmony-based approach with the pitch-based approach presented earlier. Our harmonically-generated pitch profiles contain much of the same information as the root-based profiles, so it seems somewhat redundant to use them both. Instead, we combined the melodic pitch profiles with the root profiles. Specifically, we combined the weighted melodic profile (Model 2) with the durationally-weighted and metrically-differentiated root profile (Model 8), yielding a profile with 36 values. The resulting model, Model 9, identifies the correct key on 97 of the 100 songs.

The accuracy of Model 9 may be close to the maximum possible, given that there is not 100% agreement on key labelling even among human annotators. (With regard to key, our analyses were in agreement with each other 98.2% of the time.) We inspected the three songs on which Model 9 was incorrect, to try to understand the reason for its errors. On Prince's 'When Doves Cry', the tonal centre is A but the model chose G. Much of the song consists of the progression 'Am | G | G | Am |'; thus both G and Am occur at both hypermetrically strong and weak positions. Perceptually, what favours A over G as a tonal centre seems to be the fact that Am occurs on the first measure of each group of four, while G occurs on the third; incorporating this distinction into our model might improve performance. Another song, 'California Love', consists almost entirely of a repeated one-measure progression 'F G . . |'; G is the true tonic, but F is favoured by the model due to its metrical placement, outweighing the strong support for G in the melody. Finally, in the Clash's 'London Calling', the true tonal centre is E, but the unusually prominent use of the bII chord (F major) causes G to be favoured. The verse of this song is, arguably, tonally ambiguous; what favours E over G seems to be the strong harmonic and melodic move to E at the end of the chorus.

These errors suggest various possible ways of improving the model, but there does not appear to be any single 'silver bullet' that would greatly enhance performance.

The success of our models, and in particular Model 9, points to several important conclusions about key-finding in rock. First, both the distribution of roots and the distribution of melodic pitch-classes contain useful information about key, and especially good results are obtained when these sources of information are combined. Second, the conjecture of Temperley (2001) and Stephenson (2002) regarding the metrical placement of harmonies is confirmed: key-finding from root information is considerably improved when the metrical strength of harmonies is taken into account. Finally, we see little evidence that considering chord transitions is necessary for key-finding; it appears possible to achieve nearly perfect performance by treating chords in a 'zeroth-order' fashion.

It is difficult to evaluate this model in relation to the other models of key-finding surveyed earlier. To compare symbolic key-finding models to audio ones hardly seems fair, given the greater difficulty of the latter task. The only model designed for symbolic key-finding in popular music, to our knowledge, is that of Noland and Sandler (2006), which achieved 87% on a corpus of Beatles songs (performance rose to 91% when the model was trained on each song). Their task was somewhat more difficult than ours, as their model was required to correctly identify keys as major or minor. On the other hand, our models are considerably simpler than theirs. By their own description, Noland and Sandler's model requires $2401 \times 24 = 57,624$ parameters (though it seems that this could be reduced by a factor of 12 by assuming the same parameters across all major keys and all minor keys). By contrast, our best-performing model requires just 36 parameters. In general, our experiments suggest that key-finding can be done quite effectively—at least from symbolic data—with a fairly small number of parameters.

5. Clustering songs by scale-degree distribution

Another purpose for which our corpus might be used is the categorization of songs based on musical content. While many studies have used statistical methods for the classification of popular music (for a review, see Aucouturier and Pachet (2003)), none, to our knowledge, have used symbolic representations as input. In addition, most previous studies of music classification have had practical purposes in mind (such as predicting consumer tastes). By contrast, the current study is undertaken more in the spirit of basic research: we wish to explore the extent to which rock songs fall into natural categories or clusters by virtue of their harmonic or melodic content, in the hope that this will give us a better understanding of the style.

An issue of particular interest is the validity of the distinction between major and minor keys. Corpora of popular music such as the CDM Beatles corpus (Mauch et al., 2009) and the Million-Song Dataset (Bertin-Mahieux et al., 2011) label songs as major or minor, and key-finding models that use these corpora for testing have generally adopted this assumption as well. However, a number of theorists have challenged the validity of the major/minor dichotomy for rock (Covach, 1997; Stephenson, 2002) or have proposed quite different systems for categorizing rock songs by their pitch content (Moore, 1992; Everett, 2004). As Moore has discussed, some songs are modal in construction, meaning that they use a diatonic scale but with the tonic at varying positions in the scale. For example, the Beatles' 'Let it Be', the Beatles' 'Paperback Writer', and REM's 'Losing My Religion' all use the notes of the C major scale; but the first of these three songs has a tonal centre of C (major or Ionian mode), the second has a tonal centre of G (Mixolydian mode), and the third has a tonal centre of A (natural minor or Aeolian mode). Pentatonic scales also play a prominent role in rock (Temperley, 2007a). Still other songs employ pitch collections that are neither diatonic nor pentatonic: for example, the chord progression of the chorus of the Rolling Stones' 'Jumping Jack Flash', Db major/Ab major/Eb major/Bb major, does not fit into any diatonic mode or scale. Similarly, the verse of the Beatles' 'Can't Buy me Love' features the lowered (minor) version of scale-degree 3 in the melody over a major tonic triad in the accompaniment, thus resisting classification into any conventional scale. Some authors have explained such phenomena in terms of 'blues scales', or blues-influenced inflections of diatonic or pentatonic scales (van der Merwe, 1989; Stephenson, 2002; Wagner, 2003).[2]

In short, the status of the major/minor dichotomy with regard to rock is far from a settled issue. The statistical data presented earlier in Section 3 could be seen as providing some insight in this regard. On the one hand, the fact that the scale-degree distribution of rock (both harmonic and melodic) is quite similar to that of classical music might suggest the presence of some kind of major/minor organization. On the other hand, important differences also exist: in particular, b7 is much more common than 7 overall in our melodic corpus, whereas the opposite situation is found with both major and minor keys in classical music. To better explore this topic, the following section presents ways of using our corpus to determine whether the major/minor distinction is applicable to rock, and if not, what other natural categories might exist.

5.1 The binary vector approach

To explore the presence of major and minor scales in rock, as well as other scale formations, we need some way of measuring the adherence of a song to a particular scale. As a first step, we might define 'adherence to a scale' in a strict sense, meaning that all degrees of the scale must be used and no others. To this end, we can represent the scale-degree content of each song as a binary 12-valued vector, with '1' in each

[2]A number of authors have discussed the 'blues scale', but it is rarely defined in a precise way. Van der Merwe (1989) defines it as the major scale with added flat versions of the 3rd, 5th, 6th, and 7th degrees; thus it includes all twelve chromatic degrees except b2.

position if the corresponding scale-degree occurs in the song and '0' if it does not; a song using all and only the degrees of the major scale would therefore have the vector [101011010101]. We experimented with this approach in various ways, looking at binary vectors for songs in both the melodic and harmonic corpora, as well as the 'union' vectors of the two (in which a scale-degree has the value 1 if it is present in either the melodic or harmonic analysis or both). Table 4 shows just one result: the 10 most common binary vectors from the melodic data. (Bear in mind that for six of the 200 songs, there was no melodic data.) The most common vector is the major scale, accounting for 24 of the songs; tied for sixth place is the major diatonic hexachord, scale-degrees 1−2−3−4−5−6. The other scales on the list are difficult to classify; some of these might be regarded as versions of the 'blues scale'. The second-most common vector, 1-2-b3-3-4-5-6-b7, might be called the 'pentatonic union' scale, as it is the union of the major and minor pentatonic scales (see Figure 10). The Mixolydian, Dorian, and Aeolian modes are all quite rare, accounting for two, three, and two songs respectively, suggesting that 'pure' modality is relatively infrequent in rock.

In general, we found that the 'binary-vector' approach was not very illuminating. It is difficult to make much sense of the 10 vectors in Table 4, and in any case they only account for slightly more than half of the songs in the corpus; the other songs are characterized by a large number of different vectors that occur only once or a few times. We also experimented with other kinds of binary methods of assessing adherence to scales—for example, defining a scale to be present in a song if its degrees were a subset or superset of the degrees used in the song—but these methods were also found to be unrevealing. Part of the problem is that the binary approach gives no indication of how often each scale-degree is used in a song. There might be some scale-degrees that are used only occasionally and incidentally, and do not really seem to be part of the scale of the song—just as there are in classical music (so-called 'chromatic' notes). From this point of view, it seems preferable to represent each song with a real-valued distribution of scale-degrees (either melodic or harmonic), so that the frequency of occurrence of each degree is taken into account. This is the approach we explore in the next section. While we considered combining the melodic and harmonic distributions into one, there seemed to be no principled way of doing this (what would be the weight of the melodic distribution in relation to the harmonic one?). And in any case, keeping the two distributions separate is quite revealing, as we will show.

Fig. 10. The 'pentatonic union' scale (assuming a tonic of C)

Table 4. The ten most common binary scale vectors in the melodic data.

Vector	Name of scale (if any)	Number of occurrences
101011010101	major diatonic / Ionian	24
101111010110	'pentatonic union'	18
101111110110		13
101111110111		12
101111010101		8
101011010100	major diatonic hexachord	7
101111010100		7
101111010111		6
101111011101		5
101011010111		5

5.2 Statistical clustering

To investigate the presence of scale structures in rock using the melodic and harmonic scale-degree distributions, we employed statistical clustering methods. A simple method that is well-suited to the current problem is K-means clustering (MacQueen, 1967). This approach is appropriate when the number of categories is pre-defined, and when items are located in some kind of multi-dimensional space. The procedure is as follows:

(1) Assign each item to a random category.
(2) For each category, calculate the mean position in the space of all of its members.
(3) For each item, calculate the distance between the item and each category mean, and place the item in the category whose mean is closest. Iterate over steps 2 and 3 until convergence.

In this case, the 'mean position' for a category is the mean distribution of all the songs it contains. Rather than calculating a 'distance' between a song and the category mean, we use the concept of cross-entropy: each category assigns a probability to the song's scale-degree distribution, and we assign the song to the category that assigns it highest probability. The hope is that songs with similar distributions will eventually be assigned to the same category, and that a category's distribution will be representative of the songs it contains.

The process is not guaranteed to find a global optimum, and indeed, we found it to be somewhat unstable, converging on different solutions from different initial states. To remedy this, we adjusted the procedure slightly: when assigning a song to a category, we compute the cross-entropy between all categories and all the songs they contain, categorizing the song in the way that yields the lowest total cross-entropy. Since the total distribution of the songs in a category is the same as the category distribution, and the cross-entropy between a distribution and itself is simply the entropy of the distribution, this is equivalent to finding the solution that minimizes the entropy of the category distributions. The 'goodness' of a particular solution is the mean of the entropies of the categories, with

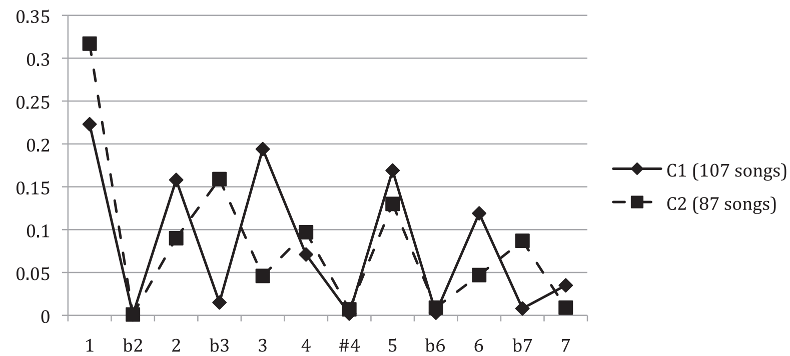

Fig. 11. Profiles for the two categories (C1 and C2) revealed by the K-means analysis of the melodic data.

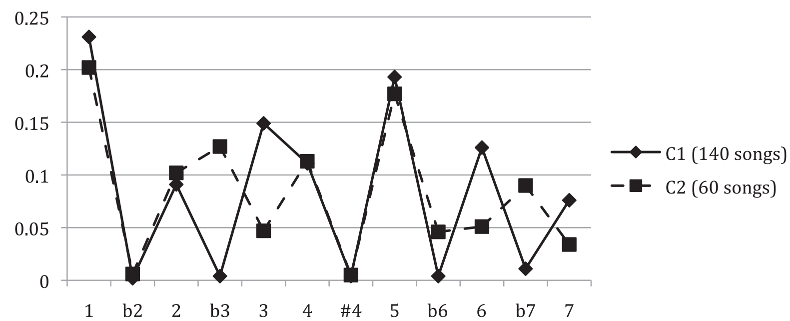

Fig. 12. Profiles for the two categories revealed by the K-means analysis of the harmonic data.

each category weighted by the number of songs it contains. This procedure was found to be stable; repeating the process from many different initial states always led to exactly the same solution.

The two-category solution for the melodic data is shown in Figure 11; the categories are arbitrarily labelled 1 and 2. Category 1 clearly represents the major scale; the seven major degrees have much higher values than the other five. Within the major scale, the five degrees of the major pentatonic scale (1–2–3–5–6) have the highest values. Category 2 is more difficult to characterize. One could call it a kind of minor distribution, since it has a much higher value for b3 than for 3. It does not, however, resemble classical minor; 6 is much more common than b6, and b7 is much more common than 7, whereas in classical minor the reverse is true in both cases (see Figure 6). It is notable also that the value for 3 is fairly high in this profile. If one were to interpret this distribution as implying a scale, the closest approximation would appear to be the 8-note scale 1–2–b3–3–4–5–6–b7; these eight degrees have far higher values than the remaining four. This is the 'pentatonic union' scale that also emerged as the second-most common binary vector in the melodic data (see Table 4 and Figure 10).

The two-category solution for the harmonic data is shown in Figure 12. Here again, category 1 very strongly reflects the major scale. Category 2 could again be described as some kind of minor (though the value for 3 is fairly high); b7 is far more common than 7 (as in the melodic 'minor' category) but 6 and b6 are nearly equal.

Both the melodic and harmonic data give some support to the validity of the major/minor distinction in rock. In both cases, two strongly distinct profiles emerge: in one profile, 3 dominates over b3, while in the other profile the reverse is true. In all four profiles, the three tonic-triad degrees (1–3–5 for category 1 profiles and 1–b3–5 for the category 2 profiles) are more common than any others, giving further support to a major/minor interpretation. It seems reasonable to refer to these profiles as 'major' and 'minor', and we will henceforth do so. While the major rock profiles strongly resemble classical major, the minor rock profiles are quite unlike classical minor, favouring the lowered seventh degree and the raised sixth, and with a significant presence of 3 as well.

Table 5. The number of songs in the major and minor harmonic and melodic categories.

Harmonic	Melodic	
	Major	Minor
Major	98	41
Minor	9	46

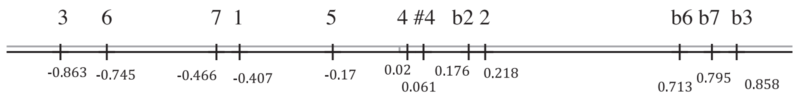

Fig. 13. The first eigenvector of the principal components analysis of the harmonic scale-degree vectors, showing the projection of each scale-degree.

In both the melodic and harmonic classification systems, there are more songs in the major category than the minor one, though the preponderance of major songs is greater in the harmonic data. Related to this, one might ask how strongly the melodic and harmonic classification schemes are correlated with one another. To address this, we can label each song by both its melodic and harmonic categories; the numbers are shown in Table 5. Not surprisingly (given the similarity between the profiles of the two categorization systems), they are indeed strongly correlated; 144 of the 194 songs (for which there is both melodic and harmonic data) are either major in both the melodic and harmonic systems, or minor in both systems. Interestingly, the vast majority of the remaining songs are minor in the melodic system and major in the harmonic one, rather than vice versa; we will call these 'minor/major' songs (stating the melodic category first). We noted earlier that some songs feature a minor third in the melody over a major tonic triad (the Beatles' 'Can't Buy me Love' was given as an example); the current data suggests that this pattern is rather common. Closer inspection of the results showed that the 'minor/major' category consists largely of 1950s songs such as Elvis Presley's 'Hound Dog', blues-influenced rock songs such as the Rolling Stones' 'Satisfaction', and soul hits such as Aretha Franklin's 'Respect'.

We also experimented with higher numbers of categories. In the melodic case, a 3-category solution (not shown here) yields one category very similar to the 2-category major profile. The other two categories both resemble the minor profile; they differ from each other mainly in that one has a much higher value for 1 and lower values for 4 and 5. (This may reflect differences in melodic *register*—the part of the scale in which the melody is concentrated—rather than in the scale itself.) In the harmonic case, the 3-category solution is quite different from the 2-category one: the first category reflects major mode, the second is similar to major but with b7 slightly higher than 7, and the third reflects Aeolian (with b3 > 3, b6 > 6 and b7 > 7). In both the melodic and harmonic data, however, the 3-category system is only very slightly lower in entropy than the 2-category one (less than 2%), thus it offers little improvement over the 2-category solution as a characterization of the data.

5.3 Principal component analysis

We explored one further approach to the classification of rock songs: principal component analysis. Let us imagine each song's scale-degree distribution (melodic or harmonic) as a point in a 12-dimensional space, with each dimension representing a scale-degree. Some dimensions may be correlated with one another, meaning that when scale-degree X has a high value in the distribution, scale-degree Y is also likely to have a high value; other dimensions may be negatively correlated. Principal component analysis searches for such correlations among dimensions and creates new axes that represent them. More precisely, it tries to establish a new coordinate system that explains as much of the variance as possible among points in the space with a small number of axes, or eigenvectors. Each eigenvector is associated with an eigenvalue, indicating the amount of variance it explains; an eigenvector that explains a large amount of variance can act as a kind of summary of statistical tendencies in the data. The analysis creates as many eigenvectors as there are dimensions in the original data, and outputs them in descending order according to how much variance they explain; the first few eigenvectors (those explaining the most variance) are generally of the most interest.

The data for the analysis is the melodic and harmonic scale-degree distributions for individual songs, used in the clustering experiment described earlier. We begin with the harmonic data. The first eigenvector (that is, the one explaining the most variance) accounted for 31% of the variance. (The second eigenvector accounted for only 19% of the variance, and we cannot find any good interpretation for it, so we will say no more about it.) Figure 13 shows the projections of each of the original dimensions on the first eigenvector. It can be seen that the three most negative dimensions are 3, 6, and 7, and the three most positive ones are b3, b6, and b7. What this tells us is that 3, 6, and 7 form a group of scale-degrees that are positively correlated with one another; b3, b6, and b7 form another group that are positively correlated; and the two groups are negatively correlated with one another. In short, the analysis provides another strong piece of evidence that major/minor dichotomy is an important dimension of variation in rock music.

A similar analysis was performed with the melodic scale-degree distributions. In this case, the first two eigenvectors were fairly close in explanatory power, the first one explaining 25% of the variance and the second one explaining 19%. The scatterplot in Figure 14 shows the projections of each scale-degree on to the first (horizontal) and second (vertical) eigenvectors. The first eigenvector appears to reflect the major/minor spectrum, though not quite as clearly as in the harmonic data; 3, 6, and 7 are positive while b3 and b7 are negative. (b6 is also negative, but only weakly so; this may simply be due to its low frequency in the melodic data.) The second eigenvector is not so easy to explain. We suspect that it reflects phenomena of register. The most positive degree in the vertical direction is 5, and the most negative one is

1; perhaps melodies tend to be centred on either 5 or 1, but not both, meaning that the two degrees are somewhat negatively correlated. Note further that #4, 4, and b6, like 5, are strongly positive; no doubt, when a melody is centred on 5, it tends to make frequent use of scale-degrees that are nearby in pitch height. This explanation only goes so far, however; 1 is strongly negative, but the degrees closest to it in height—b2, 2, b7, and 7—are not.

A small set of eigenvectors produced by a principal components analysis can be viewed as a reduction or simplification of the original space; it can then be revealing to project the original data vectors on to that reduced space. As an example, consider a hypothetical song whose harmonic scale-degree vector is a perfectly even distribution of the seven major scale-degrees, with no other scale-degrees: i.e. the vector [1/7,0,1/7,0,1/7,1/7,0,1/7,0,1/7,0,1/7]. To project this on to the eigenvector represented in Figure 13, we take the dot product of this vector with the eigenvector. This is proportional to the mean of the values of the eigenvector for which the song vector is nonzero. The location of our major ('Ionian') song on the axis is shown in Figure 15(a), as well as similar hypothetical songs in Mixolydian, Dorian, and Aeolian modes; these are the four modes that are generally said to be common in rock (Moore, 1992). It is well known that each diatonic mode contains a set of seven consecutive positions on the circle of fifths—sometimes known as the 'line of fifths' (Temperley, 2001)—and that the modes themselves reflect a natural ordering on the line (see Figure 15(b)); this is exactly the ordering that emerges from the projection in Figure 15(a). In this sense, it could be said that the line of fifths is implicit in the eigenvector (although the projection of individual scale-degrees does not reflect the line of fifths, as seen in Figure 13).[3] As noted earlier, the 'supermode'—the set of all scale-degrees commonly used in rock music, including all twelve degrees except b2 and #4—also reflects a set of adjacent positions on the line. It has been proposed that an axis of fifths plays an important role in the perception of pitch and harmony (Krumhansl, 1990) and also in the emotional connotations of melodies (Temperley & Tan, in press), suggesting that it is part of listeners' mental representation of tonal music. The implicit presence of the line of fifths in the scale-degree distribution of popular music may explain how listeners are able to internalize it from their musical experience.

[3]This argument does not work so well for the Lydian and Phrygian modes; in terms of line-of-fifths ordering of scale-degrees, Lydian should be to the left of Ionian in Figure 15(a) but is actually to its right, and Phrygian should be to the right of Aeolian but is actually to its left. Because #4 and b2 are so rare, their distribution in our corpus may not reflect their true distribution. In addition, the current analysis is assuming only a single spelling of these pitches, but in fact each one has two different spellings (#4 vs. b5, b2 vs. #1) which have different locations on the line of fifths. (For all other scale-degrees, a single spelling accounts for the vast majority of its uses.) Incorporating spelling distinctions raises many difficult practical and theoretical issues, however, so we will not attempt it here.

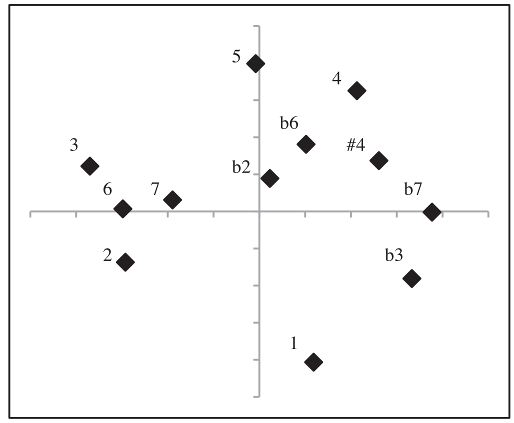

Fig. 14. The first (horizontal) and second (vertical) eigenvectors of the principal components analysis of the melodic scale-degree vectors.

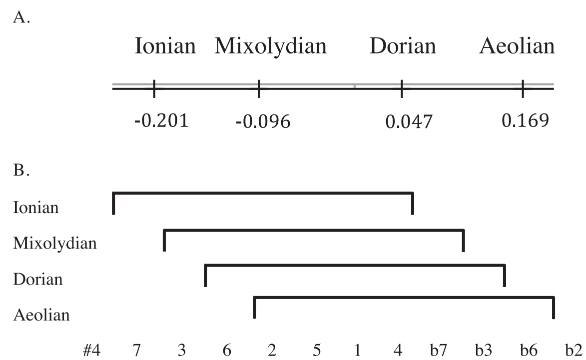

Fig. 15. (a) Projection of hypothetical modal melodies on to the axis shown in Figure 13. (b) Diatonic modes on the line of fifths.

We can also project each of the individual harmonic scale-degree vectors in our corpus on to the first harmonic eigenvector; this gives an indication of where each song lies on the major/minor axis. While we will not explore this in detail, the aggregate results are shown in Figure 16; songs are 'bucketed' into small ranges of 0.01. The distribution that emerges is of interest in two ways. First of all, it is essentially unimodal, featuring a single primary peak (there is only a very small peak toward the right end of the distribution); this suggests that rock songs are rather smoothly distributed along the major/minor dimension rather than falling into two neat categories. It is also noteworthy that the primary peak is far towards the left (major) end of the distribution; we are not sure what to make of this.

The results presented here, as well as the results of the cluster analysis presented previously, suggest that the major/minor contrast is an important dimension of variation in the pitch organization of rock music. The distinctively major scale-degrees 3, 6, and 7 tend to be used together, as do the distinctively minor degrees b3, b6, and b7; and the two groups of degrees are, to some extent at least, negatively correlated with one another. This does not mean that the major/minor system of classical tonality can be imposed wholesale on rock; indeed,

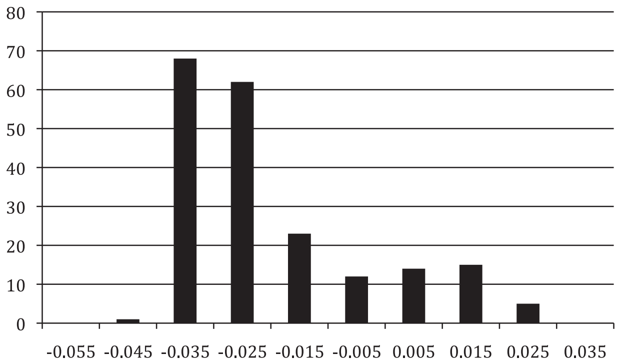

Fig. 16. Projections of the 200 harmonic scale-degree vectors on to the first harmonic eigenvector, bucketed into ranges of 0.01. Numbers on the horizontal axis indicate the centre of each bucket.

we have suggested it cannot. But some kind of major/minor spectrum is clearly operative. One might ask whether this dimension correlates in any way to conventional generic categories of rock; we suspect that it does. For example, in terms of our cluster analysis, we observe that heavy metal songs—such as Steppenwolf's 'Born to be Wild', AC/DC's 'Back in Black', and Metallica's 'Enter Sandman'—tend to be in the minor category both melodically and harmonically, while genres such as early 1960s pop (the Ronettes' 'Be My Baby', the Crystals' 'Da Doo Ron Ron') and 1970s soft rock (Elton John's 'Your Song') tend to be melodically and harmonically major. We have already noted that the 'minor/major' cluster (songs that are minor melodically and major harmonically) is associated with certain stylistic categories as well. From a practical point of view, this suggests that the major/minor dimension might be a useful predictor (in combination with others, of course) of listeners' judgments about style.

6. Conclusions and future directions

The statistical analyses and experiments presented in this paper point to several conclusions about rock. The overall scale-degree distribution of rock, both melodically and harmonically, is quite similar to that of common-practice music. In rock, as in common-practice music, all twelve degrees appear quite commonly except #4 and b2 (though b6 is borderline in the melodic distribution). The most prominent difference is that b7 is more common than 7 in rock melodies, whereas in common-practice music the reverse is true. Our experiments in key identification suggest that key-finding in rock can be done quite effectively with a purely distributional approach, using a combination of melodic scale-degree information and root information, and taking the metrical position of harmonies into account. Finally, our experiments with clustering and principal components analysis suggest that the major/minor dichotomy is an important dimension of variation in rock, though it operates quite differently from that in common-practice music in several respects. In the minor mode of rock, if one can call it that, 6 is favoured over b6, and b7 over 7; a significant proportion of songs are minor melodically but major harmonically; and the distribution of songs between major and minor appears to reflect more of a gradual continuum than two discrete categories.

One unexpected result of our analyses is the strong presence of the 'pentatonic union' scale, 1–2–b3–3–4–5–6–b7, in the melodic data. This scale is the second most common binary vector in the melodic transcriptions; it also emerges strongly in the 'minor' category of the melodic cluster analysis. Moreover, in 37 of the 194 melodies, the eight degrees of the pentatonic union scale are the *most frequent* of the twelve chromatic degrees; no other scale (of comparable size) seems to rival this except the major scale, whose degrees are the most frequent in 46 songs. Thus, three different statistical methods point to the pentatonic union scale as an important tonal structure in rock. To our knowledge, this scale has not been discussed previously. As well as being the union of the two pentatonic scales, it has several other interesting theoretical properties: it is the union of the Mixolydian and Dorian modes; it spans eight adjacent positions on the circle of fifths; and it is the union of four adjacent major triads on the circle of fifths (I-IV-bVII-bIII). The scale can also be generated by starting at 1 and moving a major second up and down (yielding b7 and 2) and a minor third up and down (adding 6 and b3), and then doing the same starting from 5 (adding 4 and 3). It could therefore be said that the scale emerges from pentatonic neighbour motion from 1 and 5; this seems significant in light of the recognized importance of stepwise melodic motion in music cognition (Bharucha, 1984; Huron, 2006).

The uses of the pentatonic union scale appear to be quite varied. In some songs, it arises from an alternation between two sections using smaller (e.g. diatonic) collections—for example, the Beatles' 'Norwegian Wood', in which the verse is Mixolydian and the bridge is Dorian (see Figure 17). In such cases, one might question whether the pentatonic union set is acting as a true 'scale'. In a number of other songs, however, the degrees of the set all appear within a single section; a case in point is the verse of the Rolling Stones' 'Satisfaction' (see Figure 18). Of particular interest here is the alternation between 3 and b3. To some extent this may be explained harmonically—the I chord in bars 1–2 of the example calls for 3 (the third of the chord), whereas the IV chord in bars 3–4 makes b3 more compatible (forming a major–minor seventh chord). But this reasoning cannot explain the b3 over the I chord at the end of the example (the beginning of the chorus). Further examination of this and other examples is needed to better understand the ways that the pentatonic union scale is used. At this point, human analysis must take over from corpus analysis; still, this example illustrates the power of corpus analysis to suggest new avenues for analytical and theoretical exploration.

The current study has barely scratched the surface of what might be done with our harmonic and melodic corpus. We have said almost nothing about rhythm, though both our melodic and harmonic corpora contain rhythmic information; this data might be used, for example, to investigate syncopation, a very important part of rock's musical language (Temperley, 2001).

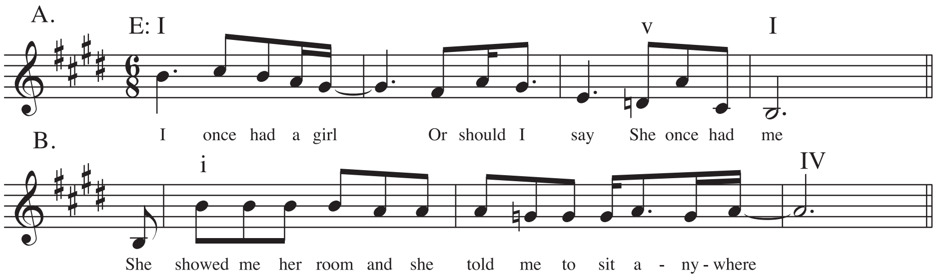

Fig. 17. The Beatles, 'Norwegian Wood'. (a) First four measures of verse; (b) first three measures of bridge.

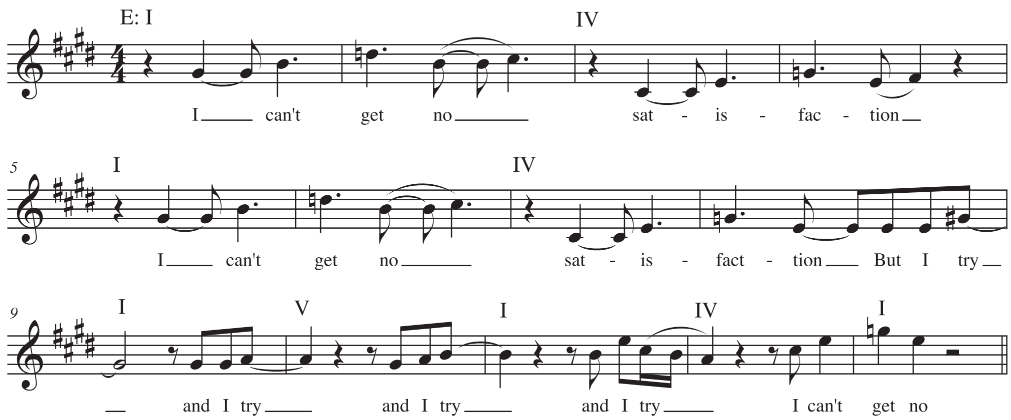

Fig. 18. The Rolling Stones, 'Satisfaction', first verse and beginning of chorus.

Much more could also be done with the pitch information in our melodic corpus. One might investigate patterns of melodic shape, as has been done quite extensively for many other musical styles (Huron, 2006). One could also examine the alignment of melody and harmony, which some authors have suggested is not as strongly constrained in rock as it is in common-practice music (Moore, 1992; Temperley, 2007a). We intend to investigate some of these issues in further work, and we hope that others will find our corpus useful as well.

References

Aucouturier, J.-J., & Pachet, F. (2003). Representing musical genre: A state of the art. *Journal of New Music Research*, *32*, 89–93.

Bertin-Mahieux, T., Ellis, D., Whitman, B., & Lamere, P. (2011). The Million Song Dataset. In Proceedings of the 12th International Society for Music Information Retrieval Conference, Miami, FL, USA.

Bharucha, J. J. (1984). Anchoring effects in music: The resolution of dissonance. *Cognitive Psychology*, *16*, 485–518.

Burgoyne, J., Wild, J., & Fujinaga, I. (2011). An expert ground truth set for audio chord recognition and music analysis. In Proceedings of the 12th International Conference on Music Information Retrieval, Miami, FL, USA.

Chai, W., & Vercoe, B. (2005). Detection of key change in classical piano music. In Proceedings of the 6th International Conference on Music Information Retrieval, London, UK.

Chew, E. (2002). The spiral array: An algorithm for determining key boundaries. In: C. Anagnostopoulou, M. Ferrand & A. Smaill (Eds.), *Music and Artificial Intelligence (pp. 18–31)*. Berlin: Springer-Verlag.

Covach, J. (1997). Progressive rock, 'Close to the Edge', and the boundaries of style. In: J. Covach & G. Boone (Eds.), *Understanding Rock (pp. 171–210)*. New York: Oxford University Press.

Cremer, M., & Derboven. C. (2004). A system for harmonic analysis of polyphonic music. In Proceedings of the 25th International Audio Engineering Society (AES) Conference, London, UK.

Dannenberg, R., Birmingham, W., Pardo, B., Hu, N., Meek, C., & Tzanetakis, G. (2007). A comparative evaluation of search techniques for query-by-humming using the MUSART testbed. *Journal of the American Society for Information Science and Technology*, *58*, 687–701.

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

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

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

Gomez, E. (2006). Tonal description of polyphonic audio for music content processing. *INFORMS Journal on Computing*, *18*, 294–304.

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

Izmirli, O. (2007). Localized key finding from audio using non-negative matrix factorization for segmentation. In Proceedings of the 8th International Conference on Music Information Retrieval, Vienna, Austria.

Kostka, S., & Payne, D. (1995). *Workbook for Tonal Harmony*. New York: McGraw-Hill.

Krumhansl, C. (1990). *Cognitive Foundations of Musical Pitch*. Oxford: Oxford University Press.

Lee, K., & Slaney, M. (2007). A unified system for chord transcription and key extraction using hidden Markov models. In Proceedings of the 8th International Conference on Music Information Retrieval, Vienna, Austria.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA.

Mauch, M., Cannam, C., Davies, M., Dixon, S., Harte, C., Kolozali, S., Tidhar, D., & Sandler, M. (2009). OMRAS2 Metadata Project (2009). *In* Proceedings of the 10th International Society for Music Information Retrieval. Kobe: Japan.

Mauch, M., & Dixon, S. (2010). Simultaneous estimation of chords and musical context from audio. *IEEE Transactions on Audio, Speech, and Language Processing*, *18*, 1280–1289.

Moore, A. (1992). Patterns of harmony. *Popular Music*, *11*, 73–106.

Moore, A. (1995). The so-called ‘flattened seventh’ in rock. *Popular Music*, *14*, 185–201.

Moore, A. (2001). *Rock: The Primary Text: Developing a Musicology of Rock*. Aldershot, UK: Ashgate Publishing.

Noland, K., & Sandler, M. (2006). Key estimation using a hidden Markov model. In Proceedings of the 7th International Conference on Music Information Retrieval, Victoria, Canada.

Papadopoulos, H., & Peeters, G. (2011). Local key estimation from an audio signal relying on harmonic and metrical structures. *IEEE Transactions on Audio, Speech, and Language Processing*, *20*(4), 1297–1312.

Patel, A., & Daniele, J. (2003). An empirical comparison of rhythm in language and music. *Cognition*, *87*, B35–B45.

Pauws, S. (2004). Musical key extraction from audio. In Proceedings of the 5th International Conference on Music Information Retrieval, Barcelona, Spain.

Purwins, H., Blankertz, B., & Obermayer, K. (2000). *A new method for trcking modulations in tonal music in audio data format*. Como, Italy: In Proceedings of the International Joint Conference on Neural Networks.

Raphael, C., & Stoddard, J. (2004). Functional harmonic analysis using probabilistic models. *Computer Music Journal*, *28*(3), 45–52.

Rocher, T., Robine, M., Hanna, P., & Oudre, L. (2010). Concurrent estimation of chords and keys from audio. In Proceedings of the 11th International Conference on Music Information Retrieval, Utrecht, Netherlands.

Rolling Stone (2004). The 500 greatest songs of all time. *Rolling Stone*, *963*, 65–165.

Schaffrath, H. (1995). In: D. Huron (Ed.), *The Essen Folksong Collection*. Stanford, CA: Center for Computer-Assisted Research in the Humanities.

Serrà, J., Gómez, E., & Herrera, P. (2010). Audio cover song identification and similarity: Background, approaches, evaluation, and beyond. In: Z. Ras & A. Wieczorkowska (Eds.), *Advances in Music Information Retrieval (pp. 307–332)*. Berlin: Springer-Verlag.

Shmulevich, I., & Yli-Harja, O. (2000). Localized key-finding: Algorithms and applications. *Music Perception*, *17*, 65–100.

Stephenson, K. (2002). *What to Listen for in Rock: A Stylistic Analysis*. New Haven: Yale University Press.

Temperley, D. (2001). *The Cognition of Basic Musical Structures*. Cambridge, MA: MIT Press.

Temperley, D. (2007a). The melodic-harmonic ‘divorce’ in rock. *Popular Music*, *26*, 323–342.

Temperley, D. (2007b). *Music and Probability*. Cambridge, MA: MIT Press.

Temperley, D., & Tan, D. (2013) Emotional connotations of diatonic modes. *Music Perception*, *30*, 237–57.

van de Laar, B. (2006). Emotion detection in music, a survey. Proceedings of the 4th Twente Student Conference on IT, Twente, Netherlands.

van der Merwe, P. (1989). *Origins of the Popular Style: The Antecedents of Twentieth-Century Popular Music*. Oxford: Clarendon Press.

von Hippel, P., & Huron, D. (2000). Why do skips precede reversals? The effect of tessitura on melodic structure. *Music Perception*, *18*, 59–85.

Vos, P. G., & Van Geenen, E. (1996). A parallel-processing key-finding model. *Music Perception*, *14*, 185–224.

Wagner, N. (2003). ‘Domestication’ of blue notes in the Beatles’ songs. *Music Theory Spectrum*, *25*, 353–65.