Arun Shenoy and Ye Wang
School of Computing
National University of Singapore
3 Science Drive 2, Singapore 117543
arun@arunshenoy.com
wangye@comp.nus.edu.sg

Key, Chord, and Rhythm Tracking of Popular Music Recordings

In this article, we propose a framework to analyze a musical audio signal (sampled from a popular music CD) and determine its key, provide usable chord transcriptions, and obtain the hierarchical rhythm structure representation comprising the quarter-note, half-note, and whole-note (or measure) levels. This framework is just one specific aspect of the broader field of content-based analysis of music. There would be many useful applications of content-based analysis of musical audio, most of which are not yet fully realized. One of these is automatic music transcription, which involves the transformation of musical audio into a symbolic representation such as MIDI or a musical score, which in principle, could then be used to recreate the musical piece (e.g., Plumbley et al. 2002). Another application lies in the field of music informational retrieval, that is, simplifying interaction with large databases of musical multimedia by annotating audio data with information that is useful for search and retrieval (e.g., Martin et al. 1998).

Two other applications are structured audio and emotion detection in music. In the case of structured audio, we are interested in transmitting sound by describing rather than compressing it (Martin et al. 1998). Here, content analysis could be used to automate partly the creation of this description by the automatic extraction of various musical constructs from the audio. Regarding emotion detection, Hevner (1936) has carried out experiments that substantiated a hypothesis that music inherently carries emotional meaning. Huron (2000) has pointed out that, because the preeminent functions of music are social and psychological, emotion could serve as a very useful measure for the characterization of music in information retrieval systems. The influence of musical chords on listeners's emotion has been demonstrated by Sollberger et al. (2003).

Whereas we would expect human listeners to be reasonably successful at general auditory scene analysis, it is still a challenge for computers to perform such tasks. Even simple human acts of cognition such as tapping the foot to the beat or swaying in time with the music are not easily reproduced by a computer program. A brief review of audio analysis as it relates to music, followed by case studies of a recently developed system that analyze specific aspects of music, has been presented by Dixon (2004). The landscape of music-content processing technologies is discussed in Aigrain (1999). The current article does not present new audio signal-processing techniques for content analysis, instead building a framework from existing techniques. However, it does represent a unique attempt at integrating harmonic and metric information within a unified system in a mutually informing manner.

Although the detection of individual notes constitutes low-level music analysis, it is often difficult for the average listener to identify them in music. Rather, it is the overall quality conveyed by the combination of notes to form chords. Chords are the harmonic description of music, and like melody and rhythm, could serve to capture the essence of the musical piece. Non-expert listeners tend to hear groups of simultaneous notes as chords. It can be quite difficult to identify whether or not a particular pitch has been heard in a chord. Furthermore, although a complete and accurate polyphonic transcription of all notes would undoubtedly yield the best results, it is often possible to classify music by genre, identify musical instruments by timbre, or segment music into sectional divisions without this low-level analysis.

Tonality is an important structural property of music, and it has been described by music theorists and psychologists as a hierarchical ordering of the pitches of the chromatic scale such that these notes are perceived in relation to one central and stable pitch, the tonic (Smith and Schmuckler 2000). This hierarchical structure is manifest in listeners's perceptions of the stability of pitches in tonal contexts. The key of a piece of music is specified by its tonic and one of two modes: major or minor. A system to determine the key of acoustic musical signals has

Computer Music Journal, 29:3, pp. 75–86, Fall 2005

been demonstrated in Shenoy et al. (2004) and will be summarized later in this article.

Rhythm is another component that is fundamental to the perception of music. Measures of music divide a piece into time-counted segments, and time patterns in music are referred to in terms of meter. The beat forms the basic unit of musical time, and in a meter of 4/4—also called common or quadruple time—there are four beats to a measure. Rhythm can be perceived as a combination of strong and weak beats. In a 4/4 measure consisting of four successive quarter notes, there is usually a strong beat on the first and third quarter notes, and a weak beat on the second and fourth (Goto and Muraoka 1994). If the strong beat constantly alternates with the weak beat, the inter-beat-interval (the temporal difference between two successive beats) would usually correspond to the temporal length of a quarter note. For our purpose, the strong and weak beats as defined above correspond to the alternating sequence of equally spaced phenomenal impulses that define the tempo of the music (Scheirer 1998). A hierarchical structure like the measure (bar-line) level can provide information more useful for modeling music at a higher level of understanding (Goto and Muraoka 1999). Key, chords, and rhythm are important expressive dimensions in musical performances. Although expression is necessarily contained in the physical features of the audio signal such as amplitudes, frequencies, and onset times, it is better understood when viewed from a higher level of abstraction, that is, in terms of musical constructs (Dixon 2003) like the ones discussed here.

Related Work

Existing work in key determination has been restricted to either the symbolic domain (MIDI and score), or, in the audio domain, single-instrument and simple polyphonic sounds (see for example, Ng et al. 1996; Chew 2001, 2002; Povel 2002; Pickens 2003; Raphael and Stoddard 2003; Zhu and Kankanhalli 2003; Zhu et al. 2004). A system to extract the musical key from classical piano sonatas sampled from compact discs has been demonstrated by Pauws (2004). Here, the spectrum is first restructured into a *chromagram* representation in which the frequencies are mapped onto a limited set of twelve chroma values. This chromagram is used in a correlative comparison with the key profiles of all the 24 Western musical keys that represent the perceived stability of each chroma within the context of a particular musical key. The key profile that has the maximum correlation with the computed chromagram is taken as the most likely key. It has been mentioned that the performance of the system on recordings from other instrumentation or from other musical idioms is unknown. Additionally, factors in music perception and cognition such as rhythm and harmony are not modeled. These issues have been addressed in Shenoy et al. (2004) in extracting the key of popular music sampled from compact-disc audio.

Most existing work in the detection and recognition of chords has similarly been restricted to the symbolic domain or single-instrument and simple polyphonic sounds (see, for example, Carreras et al. 1999; Fujishima 1999; Pardo and Birmingham 1999; Bello et al. 2000; Pardo and Birmingham 2001; Ortiz-Berenguer and Casajus-Quiros 2002; Pickens and Crawford 2002; Pickens et al. 2002; Klapuri 2003; Raphael and Stoddard 2003). A statistical approach to perform chord segmentation and recognition on real-world musical recordings that uses Hidden Markov Models (HMMs) trained using the Expectation-Maximization (EM) algorithm has been demonstrated in Sheh and Ellis (2003). This work draws on the prior idea of Fujishima (1999) who proposed a representation of audio termed "pitch-class profiles" (PCPs), in which the Fourier transform intensities are mapped to the twelve semitone classes (chroma). This system assumes that the chord sequence of an entire piece is known beforehand, which limits the technique to the detection of known chord progressions. Furthermore, because the training and testing data is restricted to the music of the Beatles, it is unclear how this system would perform for other kinds of music. A learning method similar to this has been used to for chord detection in Maddage et al. (2004). Errors in chord detection have been corrected using knowledge from the key-detection system in Shenoy et al. (2004).

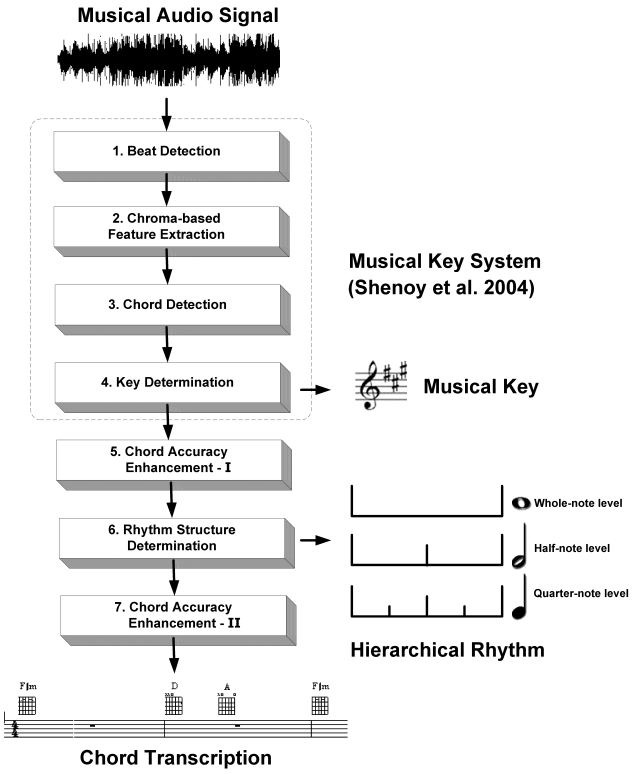
Much research in the past has also focused on rhythm analysis and the development of beat-tracking systems. However, most of this research does not consider higher-level beat structures above the quarter-note level, or it was restricted to the symbolic domain rather than working in real-world acoustic environments (see for example, Allen and Dannenberg 1990; Goto and Muraoka 1994; Vercoe 1997; Scheirer 1998; Cemgil et al. 2001; Dixon 2001; Raphael 2001; Cemgil and Kappen 2003; Dixon 2003). Goto and Muraoka (1999) perform real-time higher level rhythm determination up to the measure level in musical audio without drum sounds using onset times and chord change detection for musical decisions. The provisional beat times are a hypothesis of the quarter-note level and are inferred by an analysis of onset times. The chord-change analysis is then performed at the quarter-note level and at the interpolated eighth-note level, followed by an analysis of how much the dominant frequency components included in chord tones and their harmonic overtones change in the frequency spectrum. Musical knowledge of chord change is then applied to detect the higher-level rhythm structure at the half-note and measure (whole-note) levels. Goto has extended this work to apply to music with and without drum sounds using drum patterns in addition to onset times and chord changes discussed previously (Goto 2001). The drum pattern analysis can be performed only if the musical audio signal contains drums, and hence a technique that measures the autocorrelation of the snare drum's onset times is applied. Based on the premise that drum-sounds are noisy, the signal is determined to contain drum sounds only if this autocorrelation value is high enough. Based on the presence or absence of drum sounds, the knowledge of chord changes and/or drum patterns is selectively applied. The highest level of rhythm analysis at the measure level (whole-note/bar) is then performed using only musical knowledge of chord change patterns. In both these works, chords are not recognized by name, and thus rhythm detection has been performed using chord-change probabilities rather than actual chord information.

System Description

A well-known algorithm used to identify the key of music is the Krumhansl-Schmuckler key-finding algorithm (Krumhansl 1990). The basic principle of the algorithm is to compare the input music with a prototypical major (or minor) scale-degree profile. In other words, the distribution of pitch-classes in a piece is compared with an ideal distribution for each key. This algorithm and its variations (Huron and Parncutt 1993; Temperley 1999a, 1999b, 2002), however, could not be directly applied in our system, as these require a list of notes with note-on and note-off times, which cannot be directly extracted from polyphonic audio. Thus, the problem has been approached in Shenoy et al. (2004) at a higher level by clustering individual notes to obtain the harmonic description of the music in the form of the 24 major/minor triads. Then, based on a rule-based analysis of these chords against the chords present in the major and minor keys, the key of the song is extracted. The audio has been framed into beat-length segments to extract metadata in the form of quarter-note detection of the music. The basis for this technique is to assist in the detection of chord structure based on the musical knowledge that chords are more likely to change at beat times than at other positions (Goto and Muraoka 1999).

The beat-detection process first detects the onsets present in the music using sub-band processing (Wang et al. 2003). This technique of onset detection is based on the sub-band intensity to detect the perceptually salient percussive events in the signal. We draw on the prior ideas of beat tracking discussed in Scheirer (1998) and Dixon (2003) to determine the beat structure of the music. As a first step, all possible values of inter-onset intervals (IOIs) are computed. An IOI is defined as the time interval between any pair of onsets, not necessarily successive. Then, clusters of IOIs are identified, and a ranked set of hypothetical inter-beat-intervals (IBIs) is created based on the size of the corresponding clusters and by identifying integer relationships with other clusters. The latter process is to recognize harmonic relationships between the beat (at the quarter-note level) and simple integer multiples of the beat (at

Figure 1. System framework.



the half-note and whole-note levels). An error margin of 25 msec has been set in the IBI to account for slight variations in the tempo. The highest ranked value is returned as the IBI from which we obtain the tempo, expressed as an inverse value of the IBI. Patterns of onsets in clusters at the IBI are tracked, and beat information is interpolated into sections in which onsets corresponding to the beat might not be detected.

The audio feature for harmonic analysis is a reduced spectral representation of each beat-spaced segment of the audio based on a chroma transformation of the spectrum. This feature class represents the spectrum in terms of pitch class and forms the basis for the chromagram (Wakefield 1999). The system takes into account the chord distribution across the diatonic major scale and the three types of minor scales (natural, harmonic, and melodic). Furthermore, the system has been biased by assigning relatively higher weights to the primary chords in each key (tonic, dominant, and subdominant). This concept has also been demonstrated in the Spiral Array, a mathematical model for tonality (Chew 2001, 2002).

It is observed that the chord-recognition accuracy of the key system, though sufficient to determine the key, is not sufficient to provide usable chord transcriptions or determine the hierarchical rhythm structure across the entire duration of the music. Thus, in this article, we enhance the four-step key-determination system with three post-processing stages that allow us to perform these two tasks with greater accuracy. The block diagram of the proposed framework is shown in Figure 1.

System Components

We now discuss the three post-key processing components of the system. These consist of a first phase of chord accuracy enhancement, rhythm structure determination, and a second phase of chord accuracy enhancement.

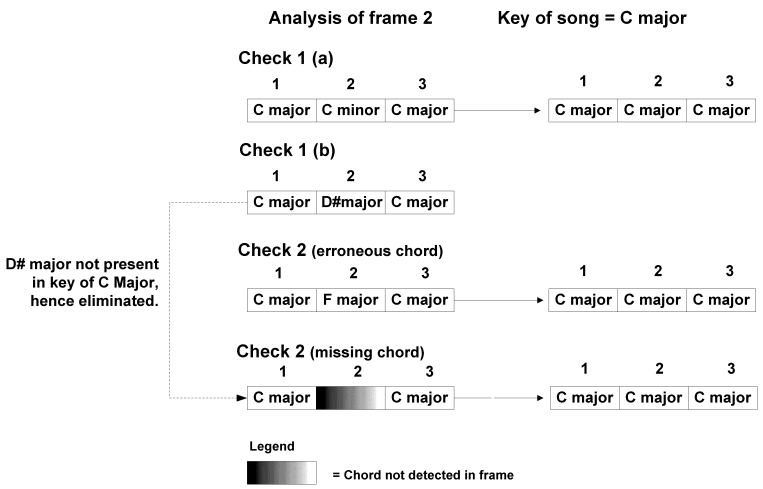
Chord Accuracy Enhancement (Phase 1)

In this step we aim to increase the accuracy of chord detection. For each audio frame, we perform two checks.

Check 1: Eliminate Chords Not in the Key of the Song

Here, we perform a rule-based analysis of the detected chord to see if it exists in the key of the song. If not, we check for the presence of the major chord of the same pitch class (if the detected chord is minor), and vice versa. If present in the key, we replace the erroneous chord with this chord. This is because the major and minor chord of a pitch class differ only in the position of their mediants. The chord-detection approach often suffers from recognition errors that result from overlaps of harmonic components of individual notes in the spectrum; these are quite difficult to avoid. Hence, there is a possibly

Figure 2. Chord accuracy enhancement (Phase 1).



of error in the distinction between the major and minor chords for a given pitch class. It must be highlighted that chords outside the key are not necessarily erroneous, and this usage is a simplification used by this system. If this check fails, we eliminate the chord.

Check 2: Perform Temporal Corrections of Detected or Missing Chords

If the chords detected in the adjacent frames are the same as each other but different from the current frame, then the chord in the current frame is likely to be incorrect. In these cases, we coerce the current frame's chord to match the one in the adjacent frames.

We present an illustrative example of the above checks over three consecutive quarter-note-spaced frames of audio in Figure 2.

Rhythm Structure Determination

Next, our system checks for the start of measures based on the premise that chords are more likely to change at the beginning of a measure than at other beat positions (Goto 2001). Because there are four quarter notes to a measure in 4/4 time (which is by far the most common meter in popular music), we check for patterns of four consecutive frames with the same chord to demarcate all possible measure boundaries. However, not all of these boundaries may be correct. We will illustrate this with an example in which a chord sustains over two measures of the music. From Figure 3c, it can be seen that there are four possible measure boundaries being detected across the twelve quarter-note-spaced frames of audio. Our aim is to eliminate the two erroneous ones, shown with a dotted line in Figure 3c, and interpolate an additional measure line at the start of the fifth frame to give us the required result as seen in Figure 3b.

The correct measure boundaries along the entire length of the song are thus determined as follows. First, obtain all possible patterns of boundary locations that are have integer relationships in multiples of four. The, select the pattern with the highest count as the one corresponding to the pattern of actual measure boundaries. Track the boundary locations in the detected pattern, and interpolate missing boundary positions across the rest of the song.

Chord Accuracy Enhancement (Phase 2)

Now that the measure boundaries have been extracted, we can increase the chord accuracy in each measure of audio with a third check.

Figure 3. Error in measure boundary detection.

Figure 4. Chord accuracy enhancement (Phase 2).

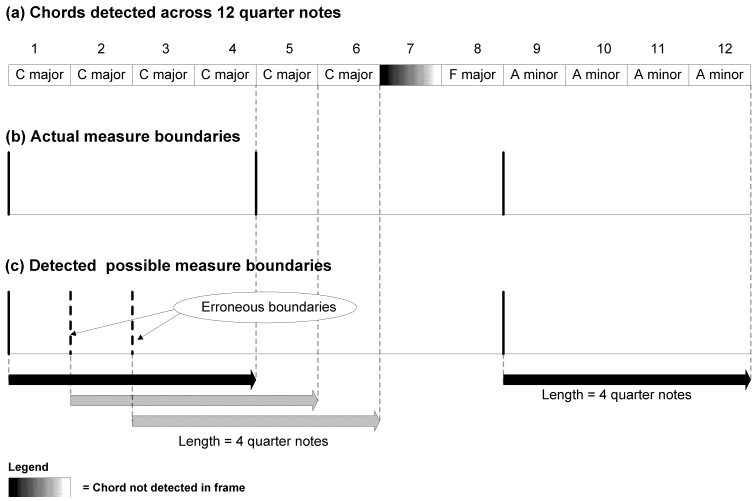


Figure 3

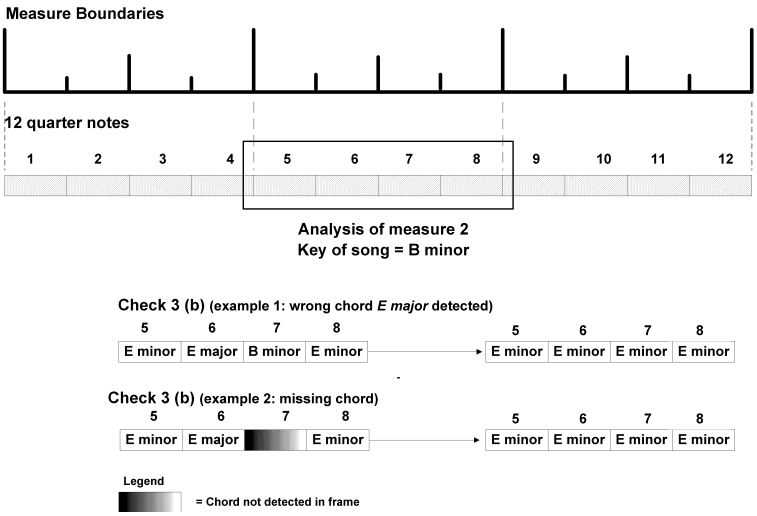


Figure 4

Check 3: Intra-Measure Chord Check

From Goto and Muraoka (1999), we know that chords are more likely to change at the beginning of the measures than at other positions of half-note times. Hence if three of the chords are the same, then the four chord is likely to be the same as the others (Check 3a). Also, if a chord is common to both halves of the measure, then all the chords in the measure are likely to be the same as this chord (Check 3b).

It is observed that all possible cases of chords under Check 3a are already handled by Checks 1 and 2 above. Hence, we only implement Check 3b, as illustrated in Figure 4 with an example. This check is required because, in the case of a minor key, we can have both the major and minor chord of the same pitch class present in the song. A classic example of

Table 1. Experimental Results

No.	*Song Title*	*Chord Detection (% accuracy)*	*Original Key*	*Detected Key*	*Chord Accuracy Enhancement-I (% accuracy)*	*Successful Measure Detection*	*Chord Accuracy Enhancement-II (% accuracy)*
1	(1965) Righteous Brothers—Unchained melody	57.68	C major	C major	70.92	28/30 songs	85.11
1	(1965) Righteous Brothers—Unchained melody	57.68	C major	C major	70.92	Yes	85.11
2	(1977) Bee Gees—Stayin' alive	39.67	F minor	F minor	54.91	Yes	71.40
3	(1977) Eric Clapton—Wonderful tonight	27.70	G major	G major	40.82	Yes	60.64
4	(1977) Fleetwood Mac—You make lovin' fun	44.37	A# major	A# major	60.69	Yes	79.31
5	(1979) Eagles—I can't tell you why	52.41	D major	D major	68.74	Yes	88.97
6	(1984) Foreigner—I want to know what love is	55.03	D# minor	D# minor	73.42	No	58.12
7	(1986) Bruce Hornsby—The way it is	59.74	G major	G major	70.32	Yes	88.50
8	(1989) Chris Rea—Road to hell	61.51	A minor	A minor	76.64	Yes	89.24
9	(1991) R.E.M.—Losing my religion	56.31	A minor	A minor	70.75	Yes	85.74
10	(1991) U2—One	56.63	C major	C major	64.82	Yes	76.63
11	(1992) Michael Jackson—Heal the world	30.44	A major	A major	51.76	Yes	68.62
12	(1993) MLTR—Someday	56.68	D major	D major	69.71	Yes	87.30
13	(1995) Coolio—Gangsta's paradise	31.75	C minor	C minor	47.94	Yes	70.79
14	(1996) Backstreet Boys—As long as you love me	48.45	C major	C major	61.97	Yes	82.82
15	(1996) Joan Osborne—One of us	46.90	A major	A major	59.30	Yes	80.05
16	(1997) Bryan Adams—Back to you	68.92	C major	C major	75.69	Yes	95.80
17	(1997) Green Day—Time of your life	54.55	G major	G major	64.58	Yes	87.77
18	(1997) Hanson—Mmmbop	39.56	A major	A major	63.39	Yes	81.08
19	(1997) Savage Garden—Truly, madly, deeply	49.06	C major	C major	63.88	Yes	80.86
20	(1997) Spice Girls—Viva forever	64.50	D# minor	F# major	74.25	Yes	91.42
21	(1997) Tina Arena—Burn	35.42	G major	G major	56.13	Yes	77.38
22	(1998) Jennifer Paige—Crush	40.37	C# min	C# min	55.41	Yes	76.78
23	(1998) Natalie Imbruglia—Torn	53.00	F major	F major	67.89	Yes	87.73
24	(1999) Santana—Smooth	54.53	A minor	A minor	69.63	No	49.91
25	(2000) Corrs—Breathless	36.77	B major	B major	63.47	Yes	77.28
26	(2000) Craig David—Walking away	68.99	A minor	C major	75.26	Yes	93.03
27	(2000) Nelly Furtado—Turn off the light	36.36	D major	D major	48.48	Yes	70.52
28	(2000) Westlife—Seasons in the sun	34.19	F# major	F# major	58.69	Yes	76.35
29	(2001) Shakira—Whenever, wherever	49.86	C# minor	C# minor	62.82	Yes	78.39
30	(2001) Train—Drops of Jupiter	32.54	C major	C major	53.73	Yes	69.85
	Overall Accuracy at each stage	**48.13**	**28/30 songs**		**63.20**	**28/30 songs**	**78.91**

this can be seen in "Hotel California" by the Eagles. This song is in the key of B Minor, and the chords in the verse include an E major and an E minor chord, which shows a possible musical shift from the ascending melodic minor to the natural minor. Here, if an E major chord is detected in a measure containing the E minor chord, Check 1 would not detect any error, because both the major and minor chords are potentially present in the key of B minor. The melodic minor, however, rarely occurs as such in popular music melodies, but it has been included in this work along with the natural and harmonic minor for completeness.

Experiments

Setup

The results of our experiments, performed on 30 popular songs in English spanning five decades of music, are tabulated in Table 1. The songs have been carefully selected to represent a variety of artists and time periods. We assume the meter to be 4/4, which is the most frequent meter of popular songs, and the tempo of the input song is assumed to be constrained between 40 and 185 quarter notes per minute.

It can be observed that the average chord-detection accuracy across the length of the entire music performed by the chord-detection step (module 3 in Figure 2) is relatively low at 48.13%. The rest of the chords are either not detected or are detected in error. This accuracy is sufficient, however, to determine the key accurately for 28 out of 30 songs in the key-detection step (module 4 in Figure 2), which reflects an accuracy of over 93% for key detection. This was verified against the information in commercially available sheet music for the songs (www.musicnotes.com, www.sheetmusicplus.com). The average chord detection accuracy of the system improves on an average by 15.07% on applying Chord Accuracy Enhancement (Phase 1). (See module 5 in Figure 2.) Errors in key determination do not have any effect on this step, as will be discussed next.

The new accuracy of 63.20% has been found to be sufficient to determine the hierarchical rhythm structure (module 6 in Figure 2) across the music for 28 out of the 30 songs, thus again reflecting an accuracy of over 93% for rhythm tracking. Finally, the application of Chord Accuracy Enhancement (Phase 2; see module 7 in Figure 2) makes a substantial performance improvement of 15.71%, leading to a final chord-detection accuracy of 78.91%. This could have been higher were it not for the performance drop for the two songs (song numbers 6 and 24 in Table 1) owing to error in measure-boundary detection. This exemplifies the importance of accurate measure detection to performing intra-measure chord checks based on the previously discussed musical knowledge of chords.

Analysis of Key

It can be observed that for two of the songs (song numbers 20 and 26 in Table 1), the key has been determined incorrectly, because the major key and its relative minor are very close. Our technique assumes that the key of the song is constant throughout the length of the song. However, many songs often use both major and minor keys, perhaps choosing a minor key for the verse and a major key for the chorus, or vice versa. Sometimes, the chords used in the song are present in both the major and its relative minor. For example, the four main chords used in the song "Viva Forever" by the Spice Girls are D-sharp minor, A-sharp minor, B major, and F-sharp major. These chords are present in the key of F-sharp major and D-sharp minor; hence, it is difficult for the system to determine if the song is in the major key or the relative minor. A similar observation can be made for the song "Walking Away" by Craig David, in which the main chords used are A minor, D minor, F major, G major, and C major. These chords are present in both the key of C major as well as in its relative minor, A minor.

The use of a weighted-cosine-similarity technique causes the shorter major-key patterns to be preferred over the longer minor-key patterns, owing to normalization that is performed while applying the cosine similarity. For the minor keys, normalization is applied taking into account the count of chords that can be constructed across all the three types of minor scales. However, from an informal evaluation of popular music, we observe that popular music in minor keys usually shifts across only two out of the three scales, primarily the natural and harmonic minor. In such cases, the normalization technique applied would cause the system to become slightly biased toward the relative major key, where this problem is not present as there is only one major scale.

Errors in key recognition do not affect the Chord Accuracy Enhancement (Phase 1) module, because we also consider chords present in the relative major/minor key in addition to the chords in the detected key. A theoretical explanation of how to perform key identification in such cases of ambiguity (as seen above) based on an analysis of sheet music can be found in Ewer (2002).

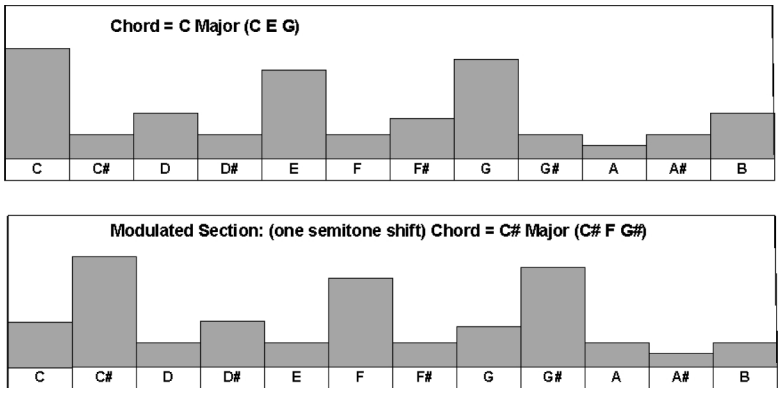
Analysis of Chord

The variation in the chord-detection accuracy of the system can be explained in terms of other chords and key change.

Use of Other Chords

In this approach, we have considered only the major and minor triads. However, in addition to these,

Figure 5. Key Shift.



there are other chord possibilities in popular music, which likely contribute to the variation in chord detection. These chords include the augmented and diminished triads, seventh chords (e.g., the dominant seventh, major seventh, and minor seventh), etc.

Polyphony, with its multidimensional sequences of overlapping tones and overlapping harmonic components of individual notes in the spectrum, might cause the elements in the chroma vector to be weighted wrongly. As a result, a Cmajor 7 chord (C, E, G, B) in the music might incorrectly get detected as an E minor chord (E, G, B) if the latter three notes are assigned a relatively higher weight in the chroma vector.

Key Change

In some songs, there is a key change toward the end of a song to make the final repeated part(s) (e.g., the chorus/refrain) slightly different from the previous parts. This is affected by transposing the song to higher semitones, usually up a half step. This has also been highlighted by Goto (2003). Because our system does not currently handle key changes, the chords detected in this section will not be recognized. This is illustrated with an example in Figure 5.

Another point to be noted here is of chord substitution/simplification of extended chords in the evaluation of our system. For simplicity, extended chords can be substituted by their respective major/minor triads. As long as the notes in the extended chord are present in the scale, and the basic triad is present, the simplification can be done. For example, the C-major 7 can be simplified to the C major triad. This substitution has been performed on the extended chords annotated in the sheet music in the evaluation of our system.

Rhythm Tracking Observations

We conjecture the error in rhythm tracking to be caused by errors in the chord-detection as discussed above. Chords present in the music and not handled by our system could be incorrectly classified into one of the 24 major/minor triads owing to complexities in polyphonic audio analysis. This can result in incorrect clusters of four chords being captured by the rhythm-detection process, resulting in an incorrect pattern of measure boundaries having the highest count. Furthermore, beat detection is a non-trivial task; the difficulties of tracking the beats in acoustic signals are discussed in Goto and Muraoka (1994). Any error in beat detection can cause a shift in the rhythm structure determined by the system.

Discussion

We have presented a framework to determine the key, chords, and hierarchical rhythm structure from acoustic musical signals. To our knowledge, this is the first attempt to use a rule-based approach that combines low-level features with high-level music knowledge of rhythm and harmonic structure to determine all three of these dimensions of music. The

methods should (and do) work reasonably well for key and chord labeling of popular music in 4/4 time. However, they would not extend very far beyond this idiom or to more rhythmically and tonally complex music. This framework has been applied successfully in various other aspects of content analysis, such as singing-voice detection (Nwe et al. 2004) and the automatic alignment of textual lyrics and musical audio (Wang et al. 2004). The human auditory system is capable of extracting rich and meaningful data from complex audio signals (Sheh and Ellis 2003), and existing computational auditory analysis systems fall clearly behind humans in performance. Towards this end, we believe that the model proposed here provides a promising platform for the future development of more sophisticated auditory models based on a better understanding of music. Our current and future research that builds on this work is highlighted below.

Key Detection

Our technique assumes that the key of the song is constant throughout the duration of the song. However, owing to the properties of the relative major/minor key combinations, we have made the chord- and rhythm-detection processes (which use the key of the song as input) quite robust against changes across this key combination. However, the same cannot be said about other kinds of key changes in the music. This is because such key changes are quite difficult to track, as there are no fixed rules and they tend to depend more on the songwriter's creativity. For example, the song "Let It Grow" by Eric Clapton switches from a B-minor key in the verse to an E major key in the chorus. We believe that an analysis of the song structure (verse, chorus, bridge, etc.) could likely serve as an input to help track this kind of key change. This problem is currently being analyzed and will be tackled in the future.

Chord Detection

In this approach, we have considered only the major and minor triads. However, in addition to these, there are other chord possibilities in popular music, and future work will be targeted toward the detection of dominant-seventh chords and extended chords as discussed earlier. Chord-detection research can be further extended to include knowledge of chord progressions based on the function of chords in their diatonic scale, which relates to the expected resolution of each chord within a key. That is, the analysis of chord progressions based on the "need" for a sounded chord to move from an "unstable" sound (a "dissonance") to a more final or "stable" sounding one (a "consonance").

Rhythm Tracking

Unlike that of Goto and Muraoka (1999), the rhythm-extraction technique employed in our current system does not perform well for "drumless" music signals, because the onset detector has been optimized to detect the onset of percussive events (Wang et al. 2003). Future effort will be aimed at extending the current work for music signals that do not contain drum sounds.

Acknowledgments

We thank the editors and two anonymous reviewers for helpful comments and suggestions on an earlier version of this article.

References

Aigrain, P. 1999. "New Applications of Content Processing of Music." *Journal of New Music Research* 28(4):271–280.

Allen, P. E., and R. B. Dannenberg. 1990. "Tracking Musical Beats in Real Time." *Proceedings of the 1990 International Computer Music Conference.* San Francisco: International Computer Music Association, pp. 140–143.

Bello, J. P., et al. 2000. "Techniques for Automatic Music Transcription." Paper presented at the 2000 International Symposium on Music Information Retrieval, University of Massachusetts at Amherst, 23–25 October.

Carreras, F., et al. 1999. "Automatic Harmonic Description of Musical Signals Using Schema-Based Chord De-

composition." *Journal of New Music Research* 28(4): 310–333.
Cemgil, A. T., et al. 2001. "On Tempo Tracking: Tempogram Representation and Kalman Filtering." *Journal of New Music Research* 29(4):259–273.
Cemgil, A. T., and B. Kappen. 2003. "Monte Carlo Methods for Tempo Tracking and Rhythm Quantization." *Journal of Artificial Intelligence Research* 18(1):45–81.
Chew, E. 2001. "Modeling Tonality: Applications to Music Cognition." *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society.* Wheat Ridge, Colorado: Cognitive Science Society, pp. 206–211.
Chew, E. 2002. "The Spiral Array: An Algorithm for Determining Key Boundaries." *Proceedings of the 2002 International Conference on Music and Artificial Intelligence.* Vienna: Springer, pp. 18–31.
Dixon, S. 2001. "Automatic Extraction of Tempo and Beat from Expressive Performances." *Journal of New Music Research* 30(1):39–58.
Dixon, S. 2003. "On the Analysis of Musical Expressions in Audio Signals." *SPIE—The International Society for Optical Engineering* 5021(2):122–132.
Dixon, S. 2004. "Analysis of Musical Content in Digital Audio." *Computer Graphics and Multimedia: Applications, Problems, and Solutions.* Hershey, Pennsylvania: Idea Group, pp. 214–235.
Ewer, G. 2002. *Easy Music Theory.* Lower Sackville, Nova Scotia: Spring Day Music Publishers.
Fujishima, T. 1999. "Real-Time Chord Recognition of Musical Sound: A System Using Common Lisp Music." *Proceedings of the 1999 International Computer Music Conference.* San Francisco: International Computer Music Association, pp. 464–467.
Goto, M. 2001. "An Audio-Based Real-Time Beat-Tracking System for Music with or without Drum Sounds." *Journal of New Music Research* 30(2):159–171.
Goto, M. 2003. "A Chorus-Section Detecting Method for Musical Audio Signals." *Proceedings of the 2003 International Conference on Acoustics, Speech, and Signal Processing.* Piscataway, New Jersey: Institute for Electrical and Electronics Engineers, pp. 437–440.
Goto, M., and Y. Muraoka. 1994. "A Beat-Tracking System for Acoustic Signals of Music." *Proceedings of the 1994 ACM Multimedia.* New York: Association for Computing Machinery, pp. 365–372.
Goto, M., and Y. Muraoka. 1999. "Real-Time Beat Tracking for Drumless Audio Signals: Chord Change Detection for Musical Decisions." *Speech Communication* 27(3–4):311–335.
Hevner, K. 1936. "Experimental Studies of the Elements of Expression in Music." *American Journal of Psychology* 48:246–268.
Huron, D. 2000. "Perceptual and Cognitive Applications in Music Information Retrieval." Paper presented at the 2000 International Symposium on Music Information Retrieval, University of Massachusetts at Amherst, 23–25 October.
Huron, D., and R. Parncutt. 1993. "An Improved Model of Tonality Perception Incorporating Pitch Salience and Echoic Memory." *Psychomusicology* 12(2): 154–171.
Klapuri, A. 2003. "Automatic Transcription of Music." *Proceedings of SMAC 2003.* Stockholm: KTH.
Krumhansl, C. 1990. *Cognitive Foundations of Musical Pitch.* Oxford: Oxford University Press.
Maddage, N. C., et al. 2004. "Content-Based Music Structure Analysis with Applications to Music Semantics Understanding." *Proceedings of 2004 ACM Multimedia.* New York: Association for Computing Machinery, pp. 112–119.
Martin, K.D., et al. 1998. "Music Content Analysis Through Models of Audition." Paper presented at the 1998 ACM Multimedia Workshop on Content Processing of Music for Multimedia Applications, Bristol, UK, 12 April.
Ng, K., et al. 1996. "Automatic Detection of Tonality Using Note Distribution." *Journal of New Music Research* 25(4):369–381.
Nwe, T. L., et al. 2004. "Singing Voice Detection in Popular Music." Paper presented at the 2004 ACM Multimedia Conference, New York, 12 October.
Ortiz-Berenguer, L. I., and F. J. Casajus-Quiros. 2002. "Polyphonic Transcription Using Piano Modeling for Spectral Pattern Recognition." *Proceedings of the 2002 Conference on Digital Audio Effects.* Hoboken, New Jersey: Wiley, pp. 45–50.
Pardo, B., and W. Birmingham. 1999. "Automated Partitioning of Tonal Music." Technical Report CSE-TR-396–99, Electrical Engineering and Computer Science Department, University of Michigan.
Pardo, B., and W. Birmingham. 2001. "Chordal Analysis of Tonal Music." Technical Report CSE-TR-439-01, Electrical Engineering and Computer Science Department, University of Michigan.
Pauws, S. 2004. "Musical Key Extraction from Audio." Paper presented at the 2004 International Symposium on Music Information Retrieval, Barcelona, 11 October.
Pickens, J., et al. 2002. "Polyphonic Score Retrieval Using Polyphonic Audio Queries: A Harmonic Modeling Approach." Paper presented at the 2002 International

Symposium on Music Information Retrieval, Paris, 15 October.
Pickens, J., and T. Crawford. 2002. "Harmonic Models for Polyphonic Music Retrieval." *Proceedings of the 2002 ACM Conference on Information and Knowledge Management.* New York: Association for Computing Machinery, pp. 430–437.
Pickens, J. 2003. "Key-Specific Shrinkage Techniques for Harmonic Models." Poster presented at the 2003 International Symposium on Music Information Retrieval, Baltimore, 26–30 October.
Plumbley, M. D., et al. 2002. "Automatic Music Transcription and Audio Source Separation." *Cybernetics and Systems* 33(6):603–627.
Povel, D. J. L. 2002. "A Model for the Perception of Tonal Melodies." *Proceedings of the 2002 International Conference on Music and Artificial Intelligence.* Vienna: Springer, pp. 144–154.
Raphael, C. 2001. "Automated Rhythm Transcription." Paper presented at the 2001 International Symposium on Music Information Retrieval, Bloomington, Indiana, 15–17 October.
Raphael, C., and J. Stoddard. 2003. "Harmonic Analysis with Probabilistic Graphical Models." Paper presented at the 2003 International Symposium on Music Information Retrieval, Baltimore, 26–30 October.
Scheirer, E. D. 1998. "Tempo and Beat Analysis of Acoustic Musical Signals." *Journal of the Acoustical Society of America* 103(1):588–601.
Sheh, A., and D. P. W. Ellis. 2003. "Chord Segmentation and Recognition Using Em-Trained Hidden Markov Models." Paper presented at the 2003 International Symposium on Music Information Retrieval, Baltimore, 26–30 October.
Shenoy, A., et al. 2004. "Key Determination of Acoustic Musical Signals." Paper presented at the 2004 International Conference on Multimedia and Expo, Taipei, 27–30 June.
Smith, N. A., and M. A. Schmuckler. 2000. "Pitch-Distributional Effects on the Perception of Tonality." Paper presented at the Sixth International Conference on Music Perception and Cognition, Keele, 5–10 August.
Sollberger, B., et al. 2003. "Musical Chords as Affective Priming Context in a Word-Evaluation Task." *Music Perception* 20(3):263–282.
Temperley, D. 1999a. "Improving the Krumhansl-Schmuckler Key-Finding Algorithm." Paper presented at the 22nd Annual Meeting of the Society for Music Theory, Atlanta, 12 November.
Temperley, D. 1999b. "What's Key for Key? The Krumhansl-Schmuckler Key-Finding Algorithm Reconsidered." *Music Perception* 17(1):65–100.
Temperley, D. 2002. "A Bayesian Model of Key-Finding." Proceedings of the 2002 International Conference on Music and Artificial Intelligence. Vienna: Springer, pp. 195–206.
Vercoe, B. L. 1997. "Computational Auditory Pathways to Music Understanding." In I. Deliège and J. Sloboda, eds. *Perception and Cognition of Music.* London: Psychology Press, pp. 307–326.
Wakefield, G. H. 1999. "Mathematical Representation of Joint Time-Chroma Distributions." *SPIE—The International Society for Optical Engineering* 3807:637–645.
Wang, Y., et al. 2003. "Parametric Vector Quantization for Coding Percussive Sounds in Music." *Proceedings of the 2003 International Conference on Acoustics, Speech, and Signal Processing.* Piscataway, New Jersey: Institute for Electrical and Electronics Engineers.
Wang, Y. et al. 2004. "LyricAlly: Automatic Synchronization of Acoustic Musical Signals and Textual Lyrics." *Proceedings of 2004 ACM Multimedia.* New York: Association for Computing Machinery, pp. 212–219.
Zhu, Y., and M. Kankanhalli. 2003. "Music Scale Modeling for Melody Matching." *Proceedings of the Eleventh International ACM Conference on Multimedia.* New York: Association for Computing Machinery, pp. 359–362.
Zhu, Y., et al. 2004. "A Method for Solmization of Melody." Paper presented at the 2004 International Conference on Multimedia and Expo, Taipei, 27–30 June.