Key Determination of Acoustic Musical Signals

Arun Shenoy, Roshni Mohapatra, Ye Wang
School of Computing, National University of Singapore, Singapore 117543
{arunshen, roshnimo, wangye} @comp.nus.edu.sg

Abstract

This paper presents a novel rule-based approach for determining the key of acoustic musical signals. Knowledge of the key will enable us to derive from music knowledge the pitch class elements that a piece of music uses. Our technique is a combination of Chroma based frequency analysis and music knowledge of rhythm structure and chord change patterns followed by rule-based inference. Experimental results illustrate that 90% accuracy is achieved for key determination and we have given an explanation for the cases that have generated an incorrect key. Steps for further development are also outlined.

1. Introduction

A computational model that can understand musical audio signals in a human-like fashion is one of the goals of computational auditory scene analysis [1]. A popular approach to this goal is to build an automatic music transcription system which typically transforms audio signals into a symbolic representation such as a musical score. However, a computational transcription system falls clearly behind humans in performance and is an actively researched area. So towards this end, we first build a computer system that can analyze a small segment of a given music signal and determine the key of the music. The key defines the diatonic scale which a piece of music uses. The diatonic scale is a seven note scale and is most familiar as the Major scale or the Minor scale in music. In addition to serving as an input to a computational music transcription system, the key can be used to obtain high level information about the musical content of the song that can capture much of the character of the musical piece.

Our technique aims to categorize a given music signal into one of the 12 Major or 12 Minor keys. From a theoretical perspective, identification of the frequency components corresponding to the 12 pitch class elements followed by a rule based analysis would enable us to determine the key of a song. However, overlap of harmonic components of individual notes and various kinds of noise in real-world musical recordings would make this a difficult task. So, our technique aims to detect the occurrence of chords in the music signal and then apply music knowledge to determine the key of the song.

Over the years, considerable work has been done in the detection and recognition of chords. However this has been mostly restricted to single instrument sounds or simple polyphonic sounds and would be unsuitable for complex signals such as ensemble performances. Chord segmentation and recognition on real-world musical recordings using Expectation-Maximization (EM) trained Hidden Markov Models (HMMs) has been used in [2] by Sheh and Ellis. This approach does not incorporate higher level musical knowledge and the training approach limits the technique to be restricted to the detection of known chord progressions. In [1], Goto and Muraoka have developed a technique for detecting a hierarchical beat structure in musical audio without drum-sounds using chord change detection for musical decisions. This technique uses musical knowledge to perform analysis over the quarter note and eighth note levels, but detects only changes in the chord progression without requiring the chord to be identified. In [3], Goto has developed a hierarchical beat tracking system for music with or without drum sounds using a combination of onset times, chord changes and drum patterns. Here again, the system does not identify musical notes or chords by name.

We draw on the prior idea of Goto and Muraoka in [1,3] to incorporate higher level music knowledge of the relation between rhythm and chord change patterns. Our technique is based on a combination of bottom-up and top-down approaches, combining the strength of low-level features and high-level musical knowledge.

2. System

2.1. Rhythm Extraction

A combination of strong and weak beats can be perceived as the sensation of rhythm. A strong beat usually corresponds to the first and third quarter note in a measure and the weak beat corresponds to the second and fourth quarter note in a measure [4]. If the strong beat constantly alternates with the weak beat, the inter-beat-interval (the temporal difference between two successive beats) corresponds to the length of a quarter note. For our purpose, the strong and weak beat as defined above, corresponds to an alternating sequence of equally spaced phenomenal impulses which define the tempo for the music [5]. We assume the meter to be 4/4, this being the most frequent meter of popular songs and the tempo of the input song to be constrained between 40-185 M.M. (Mälzel's Metronome: the number of quarter notes per minute) and almost constant.

The musical signal is framed into beat-length segments to extract metadata in the form of quarter note detection of the music. The basis for this technique of audio framing is to assist in the detection of chord structures in the music based on the music knowledge that chords are more likely to change on beat times than on other positions [3]. Our beat extraction process first detects the onsets present in the music using sub-band processing [6]. This technique of onset detection is based on the sub-band intensity to detect the perceptually salient percussive events in the music signal. In the next step a dynamic programming approach is employed to check for patterns of equally spaced strong and weak beats among the onsets detected to determine the rhythm structure of the music as shown in Figure 1.

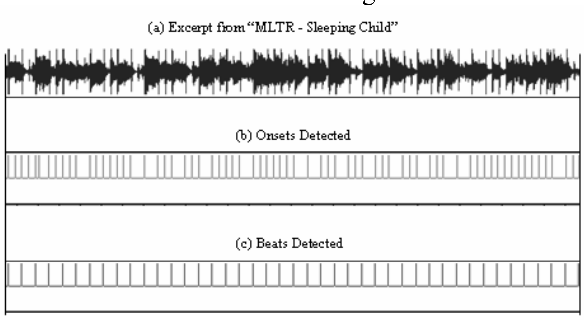

Figure 1. Rhythm Extraction

2.2 Chroma Based Feature Extraction

As highlighted in [7], there are two distinct attributes of pitch perception, Tone Height and Chroma [8]. Tone Height describes the general increase in the pitch of a sound as its frequency increases. Chroma, on the other hand, is cyclic in nature with octave periodicity. Chroma is closely related to the theoretical concept of pitch class. Under this formulation two tones separated by an integral number of octaves share the same value of Chroma. Later, it has been suggested that one could decompose frequency into similar attributes [9]. The feature which we are using 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 [10]. We restructure the frequency spectrum over 5 octaves (C2 [65.406Hz] – B6 [1975.533Hz]) into a Chroma spectrum. This mapping procedure provides us with a highly reduced representation of the frame, consisting of a single 12-element feature vector corresponding to the 12 pitch classes. We have found it useful to employ the musical relevance of Chroma in the development of features for our purpose since various 3-element pitch class combinations in the Chroma vector can be used to detect the presence of Major and Minor chords in an audio frame as will be explained in the following section.

2.3. Chord Detection

Table 1 lists the notes that are present in the Major and Minor Scales for the C pitch class.

Table 1. Pitch notes in Major/Minor scales of C

Scale	Notes in Scale							
	I	II	III	IV	V	VI	VII	I
Major	C	D	E	F	G	A	B	C
Natural Minor	C	D	D♯	F	G	G♯	A♯	C
Harmonic Minor	C	D	D♯	F	G	G♯	B	C
Melodic Minor	C	D	D♯	F	G	A	B	C

There is only one Major scale and three types of Minor scales for each of the 12 pitch classes. The Major scale follows a pattern of: "*T-T-S-T-T-T-S*" and the Natural Minor scale follows a pattern of "*T-S-T-T-S-T-T*" where *T* (implying Tone/Whole-Step in music theory) and *S* (implying Semitone/Half-Step in music theory) corresponds to a jump of one and two pitch classes respectively. The Harmonic Minor scale is obtained by raising the VII note in the Natural Minor Scale by one pitch class and the Melodic Minor scale is obtained by raising the VI note in addition to the VII note by one pitch class. The Major chords are formed by the combination of the I, III and V note of the Major Scale and the Minor chords are formed by the combination of the I, III and V note of the Minor Scale. These 3 notes correspond to the *Root*, *Mediant* and *Dominant* notes in the scale respectively. The Major and Minor chord for a pitch class differ only in the position of the Mediant note. For a Minor chord, it is one pitch class lower than the one for the Major chord. For example, the C Major chord is comprised of C, E and G notes and the C Minor chord is comprised of C, D# and G notes. For our analysis we consider only the elements with the four highest values in the Chroma vector and assign weights to them accordingly. Four elements are sufficient to distinguish between a Major and Minor chord for a given pitch key since the Major and Minor chords are *triads* (chords that are composed of 3 notes) and they differ only in the position of the Mediant note as explained previously. The chords detected across all the frames are then populated into a 24 element vector whose elements correspond to the 12 Major chords and the 12 Minor chords. The system of musical chord recognition in a polyphonic musical signal has been approached as a combination of spectral analysis to identify the individual notes followed by symbolic inference to determine the chord. However this approach often suffers from recognition errors that result from overlapping harmonic components of individual notes in the spectrum and various kinds of noise that is quite difficult to avoid.

2.4. Key Determination

Table 2 lists the chords patterns that can be present in the Major and Minor scales for C.

Table 2: Chords in C Major and C Minor keys

Scale	Chords							
	I	II	III	IV	V	VI	VII	I
Major	C maj	D min	E min	F maj	G maj	A min	B dim	C maj
Natural Minor	C min	D dim	D♯ maj	F min	G min	G♯ maj	A♯ maj	C min
Harmonic Minor	C min	D dim	D♯ aug	F min	G maj	G♯ maj	B dim	C min
Melodic Minor	C min	D min	D♯ aug	F maj	G maj	A dim	B dim	C min

This is obtained by a theoretical distribution of various chord structures (Major, Minor, Diminished and Augmented chords) to the notes in the scale (see Table 1) and the same pattern applies to all the other pitch classes. We have considered only the Major and Minor chords since they constitute the majority of the chords for any key and have found them to be sufficient to determine the key of the song. Our system assumes that the key of the song is constant throughout the length of the song. There are only 2 kinds of keys possible: Major and Minor. The chord patterns in the 3 Minor scales listed are all clustered together as being simply the Minor key. Since every song is in a particular Major or Minor key, it can be seen that there are a lot more chord choices for songs in Minor keys than for songs in a Major key. This makes the analysis of songs in a Minor key more difficult than for those in Major keys. The 24 element vector of Major and Minor chords formerly created is then pattern matched using weighted Cosine Similarity against 24 element reference vectors created for the 12 Major and 12 Minor keys. The pattern that returns the highest rank is selected as the one being the key of the song. In a given key, there are three important chords known as *primary chords* [11]. The first is the triad built on the root or tonic note, and it is called the *root* or the *tonic* chord. The next is the chord built on the fifth note, called the *dominant* chord. The third chord is built on the fourth note, and is called the *subdominant*. In the key of C Major these chords are C Major, G Major and F Major respectively. This knowledge has been incorporated in our approach by assigning relatively higher weights to the primary chords in each key in the similarity analysis.

3. Experimental Results

Table 3: Experimental Results

Song	Original key	Detected key
(1971) Don McLean - American Pie	G maj	G maj
(1976) ABBA - Dancing Queen	A maj	A maj
(1976) Eagles - Hotel California	B min	B min
(1977) Eric Clapton - Wonderful Tonight	G maj	G maj
(1977) Bee Gees - Stayin' Alive	F min	F min
(1984) Foreigner - I want to know what love is	D♯ min	D♯ min
(1986) Roxette - It must have been love	C maj	C maj
(1989) Richard Marx - Angelia	C min	D♯ maj
(1989) Chris Rea - Road to hell	A min	A min
(1992) Michael Jackson - Heal the World	A maj	A maj
(1993) MLTR – Sleeping Child	D maj	D maj
(1993) Sting - Fields of Gold	B min	B min
(1993) Michael Bolton - Said I Loved You...But I Lied	D♯ maj	D♯ maj
(1996) Backstreet Boys - As Long As you Love me	C maj	C maj
(1997) Bryan Adams – Back To You	C maj	C maj
(1997) Spice Girls – Viva Forever	D♯ min	F♯ maj
(1998) Jennifer Paige - Crush	C♯ min	C♯ min
(1998) Natalie Imbruglia – Torn	F maj	F maj
(1999) Santana - Smooth	A min	A min
(2002) Avril Lavigne - Complicated	F maj	F maj

The results of our experiments, performed on 20 popular English songs spanning 4 decades of music are tabulated in Table 3. We have managed to determine the key accurately for 18 out of the 20 songs. The original key has been obtained from commercially available sheet music for the songs. Tests performed for different song lengths have shown us that analysis over 16 bars of audio (i.e. 64 beat spaced segments of the audio) is sufficient to determine the key of the song. It can be seen here that for the song "*Richard Marx– Angelia*", the key has been wrongly determined as D# Major instead of C Minor. The explanation for this can be based on the theory of the *Relative* Major/Minor key that exists for every pitch key. Every Major scale has a relative Minor scale. The two scales are built from the exact same notes and the only difference between them is which note the scale starts with. The relative Minor scale starts from the sixth note of the Major scale. For example, the C Major scale is made up of the notes: "*C-D-E-F-G-A-B-C*" and its relative Minor scale, which is A Minor is made up of the notes "*A-B-C-D-E-F-G-A*". A Minor is called the relative Minor of C Major, and C Major is the relative Major of A Minor.

The relative Major/Minor key combination for all the 12 pitch classes is illustrated in Table 4.

Table 4: Relative Major and Minor Keys

Major	C	C♯	D	D♯	E	F	F♯	G	G♯	A	A♯	B
Minor	A	A♯	B	C	C♯	D	D♯	E	F	F♯	G	G♯

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. This has a nice effect, as it helps break up the monotony that sometimes results when a song lingers in one key. Often, when switching to a Major key from a Minor key, the songwriters will choose to go to the Relative Major from the Minor key the song is in and vice-versa. So, for example, if a song is in the key of C Minor, the relative major of that key would be D# Major. This can be taken as a probable explanation for the result above which also holds true for the second erroneous result for the song "*Spice Girls – Viva Forever*" where the actual key (D# Minor) is the relative minor of the wrongly detected key (F# Major). The usage of weighted Cosine similarity technique causes the shorter Major key patterns to be preferred over the longer Minor key patterns in the key pattern matching step.

4. Conclusion

We have presented a technique for the extraction of the key 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 its relation to chord change patterns to identify the key of the music signal. Our technique has managed to achieve 90% accuracy for tests conducted on 20 popular songs spanning 4 decades of music.

The chord recognition accuracy of the system is not yet sufficient to provide usable chord transcriptions. In this approach we have considered only the Major and Minor chords to determine the key. However, in addition to Majors and Minors, there are other chord possibilities in popular music as highlighted in [12]. Future work would be targeted towards the detection of these chords along with making the chord detection process more robust. This is with the aim of achieving accurate chord detection in every frame of audio leading to the time-measure distribution of the chords across the entire length of the music signal. The rhythm extraction technique employed in our current system does not perform very well for drumless music signals since the onset detector has been optimized to detect the onset of percussive events. Future effort will be aimed at extending the current work for music signals that do not contain drum sounds.

5. References

[1] Goto, M., and Muraoka, Y., "Real-time Beat Tracking for Drumless Audio Signals: Chord Change Detection for Musical Decisions", Speech Communication, Vol.27, Nos.3-4, April 1999.

[2] Sheh, A. and Ellis, D., "Chord Segmentation and Recognition using EM-Trained Hidden Markov Models", ISMIR 2003.

[3] Goto, M., "An Audio-based Real-time Beat Tracking System for Music With or Without Drum-sounds", JNMR, Vol.30, No.2, June 2001.

[4] Goto, M., and Muraoka, Y., "A Beat Tracking System for Acoustic Signals of Music", ACM-MM, 1994.

[5] Scheirer, E.D., "Tempo and Beat Analysis of Acoustic Musical Signals", JASA,103(1), January 1998

[6] Wang, Y., Tang, J., Ahmaniemi, A., Vaalgamaa, M., "Parametric Vector Quantization For Coding Percussive Sounds in Music", ICASSP, 2003, IEEE

[7] Bartsch, M. and Wakefield, G.H., "To Catch a Chorus: Using Chroma-Based Representations For Audio Thumbnailing", WASPAA 2001, IEEE.

[8] R. Shepard, "Circularity in judgments of relative pitch", JASA, vol. 36, 1964.

[9] Patterson, R. D., "Spiral detection of periodicity and the spiral form of musical scale", in Psychology of Music, chapter 14, 1986.

[10] Wakefield, G.H., "Mathematical representation of joint time-chroma distributions", in SPIE, 1999.

[11] Rameau, J.P., "Treatise on Harmony", Dover Publications, reprint edition, December 1984

[12] Lynn, P., "The Functions of Chords: For Pop, Jazz, and Modern Styles", P&R Press, September 2002.