AUTOMATIC KEY PARTITION BASED ON TONAL ORGANIZATION INFORMATION OF CLASSICAL MUSIC

Lam Wang Kong, Tan Lee
Department of Electronic Engineering
The Chinese University of Hong Kong Hong Kong SAR, China
{wklam,tanlee}@ee.cuhk.edu.hk

ABSTRACT

Key information is a useful information for tonal music analysis. It is related to chord progressions, which follows some specific structures and rules. In this paper, we describe a generative account of chord progression consisting of phrase-structure grammar rules proposed by Martin Rohrmeier. With some modifications, these rules can be used to partition a chord symbol sequence into different key areas, if modulation occurs. Exploiting tonal grammar rules, the most musically sensible key partition of chord sequence is derived. Some examples of classical music excerpts are evaluated. This rule-based system is compared against another system which is based on dynamic programming of harmonic-hierarchy information. Using Kostka-Payne corpus as testing data, the experimental result shows that our system is better in terms of key detection accuracy.

1. INTRODUCTION

Chord progression is the foundation of harmony in tonal music and it can determine the key. The *key* involves certain melodic tendencies and harmonic relations that maintain the tonic as the centre of attention [4]. Key is an indicator of the musical style or character. For example, the key C major is related to innocence and pureness, whereas F minor is related to depression or funereal lament [16]. Key detection is useful for music analysis. A classical music piece may have several modulations (key changes). A change of key means a change of tonal center, the adoption of a different tone to which all the other tones are to be related [10]. Key change allows tonal music to convey a sense of long-range motion and drama [17].

Keys and chord labels are interdependent. Even if the chord labels are free from errors, obtaining the key path is often a non-trivial task. For example, if a music excerpt has been analyzed with the chord sequence $[B\flat, F, G_{min}, A_{min}, G, C]$, how would you analyze its key? Is it a phrase entirely in B♭ major or C major, as they are the beginning or ending chords? Seems it is not, as B♭ major chord is normally not a member chord of C major and vice versa. It seems that there must be a key change in the middle. But how would you find out the point of key change, and how does the key change? With the help of the tonal grammar tree analysis in §2.1, a good estimate of the key path can be obtained. To start with, we assume that the excerpt consists of harmonically complete phrase(s) and the chord labels are free from errors.

There are some existing algorithms to estimate the key based on chord progression. These algorithms can be classified into two categories: *statistical-based* and *rule-based* approach. Hidden Markov model is very often used in the statistical approach. Lee & Stanley [7] extracted key information by performing harmonic analysis on symbolic training data and estimated the model parameters from them. They built 24 key-specific HMMs (all major and minor keys) for recognizing a single global key which has the highest likelihood. Raphael & Stoddard [11] performed harmonic analysis on pitch and rhythm. They divided the music into a fixed musical period, usually a measure, and associate a key and chord to each of period. They performed functional analysis of chord progression to determine the key. Unlabeled MIDI files were used to train the transition and output distributions of HMM. Instead of recognizing the global key, it can track the local key. Catteau *et al.* [2] described a probabilistic framework for simultaneous chord and key recognition. Instead of using training data, Lerdahl's representation of tonal space [8] were used as a distance metric to model the key and chord transition probabilities. Shenoy *et al.* [15] proposed a rule-based approach for determining the key from chord sequence. They created a reference vector for each of the 12 major and minor keys, including the possible chords within the key. Higher weights were assigned to primary chords (tonic, subdominant and dominant chords). The chord vector obtained from audio data were compared against the reference vector using weighted cosine similarity. The pattern with the highest rank is chosen as the selected global key.

This paper uses a rule-based approach to model tonal harmony. A context-free dependency structure is used to exhaust all the possible combinations of key paths, and the best one is selected according to music knowledge. The main objective of this research is to exploit this tonal context-free dependency structure in order to partition an excerpt of classical music into several key sections.

Functional level		Scale degree level
$TR \to DR\, T$		$T \to I$
$DR \to SR\, D$		$T \to I\; IV\; I$
$TR \to TR\; DR$		$S \to IV$
$XR \to XR\; XR$		$D \to V \mid vii$
$phrase \to TR$		$T \to vi \mid III$
		$D \to VII$ (minor)
Added rules for scale degree level:		$S \to ii$ (major)
$S \to ii$ (minor)		$S \to VI \mid bII$ (minor)
$T \to I\; IV\; VI \mid VI\; IV\; I \mid I\; bII\; I$		$X \to D(X)\; X$
$D \to I\; V$, after S or $D(V)$		$D(X) \to V/X \mid vii/X$

TR	tonic region		S	predominant function
DR	dominant region		X	any specific function
SR	predominant region		$D(\cdot)$	secondary dominant
XR	any specific region		X / Y	X of Y chord
T	tonic function		*I, III...*	major chords
D	dominant function		*ii, vi...*	minor chords

Table 1. Rules (top) and labels (bottom) used in our system

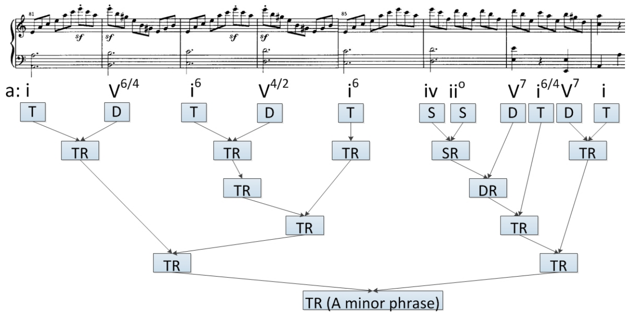

Figure 1. Example of a tonal grammar tree (single key)

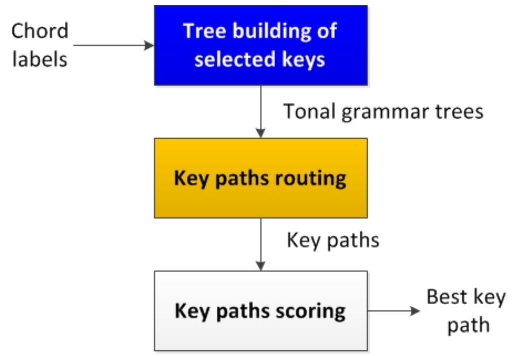

Figure 3. Flow diagram of our key partitioning system

2. TONAL THEORY OF CLASSICAL MUSIC

2.1 Schenkerian analysis and formalization

To interpret the structure of the tonal music, Schenkerian analysis [14] is used. The input is assumed to be classical music with one or more tonal centre (tonal region). Each tonal centre can be elaborated into tonic – dominant – tonic regions [1]. The dominant region can be further elaborated into predominant-dominant regions. Each region can be recursively elaborated to form a *tonal grammar tree*. We can derive the key information by referring to the top of the tree, which groups the chord sequence into a tonal region.

Context-free grammar can be used to formalize this tree structure. A list of generative syntax is proposed by Rohrmeier [13] in the form of $V \to w$. V is a single non-terminal symbol, while w is a string of terminals and/or non-terminals. Chord symbols (eg. IV) are represented by terminals. They are the leaves of the grammar tree. Tonal functions (eg. T for tonic) or regions (eg. TR for tonic region) are represented by non-terminals. They can be the internal nodes or the root of the grammar tree. For instance, the rule $D \to V \mid vii$ indicates that the V or vii chord can be represented by the dominant function. The rule $S \to ii$ *(major)* indicates that *ii* chord can be represented by the predominant function only when the current key is major. Originally Rohrmeier has proposed 28 rules. Some of them were modified to suit classical music and were listed in Table 1.

Based on this set of rules, Cocke–Younger–Kasami parsing algorithm [18] is used to construct a tonal grammar tree. If a music input is harmonically valid, a single tonal grammar tree can be built like in Figure 1. Else some scattered tree branches are resulted and cannot be connected to one single root.

2.2 Modulation

In Rohrmeier's generative syntax of tonal harmony, modulation is formalized as a new local tonic [13]. Each functional region (new key section) is grouped as a single non-tonic chord in the original passage, and they may relate this (elaborated) chord to the neighbouring chords.

In this research we have a more general view of modulation. As a music theorist, Reger had published a book *Modulation*, showing how to modulate from C major / minor to every other key [12]. Modulation to every other key is possible, but modulation to harmonically closer keys is more common [10]. For instance, if the music is originally in C major, it is more probable to modulate to G major instead of B major. Lerdahl's chordal distance [8] is used to measure the distance between different keys. Here Rohrmeier's modulation rules in [13] are not used. Instead, a tonal grammar tree is built for each new key section, and the key path with the best score is chosen. Any key changes explainable by tonicization (temporary borrowing of chords from other keys), such as the chords *[I **V/V** V I]*, is not considered as a modulation. Figure 2 shows an example of tonal grammar tree with modulation, from E minor to D♯ minor. It is presented by two disjunct trees.

3. SYSTEM BUILDING BLOCKS

3.1 Overview

The proposed key partitioning system is shown as in Figure 3. This system takes a sequence of chord labels (e.g. A minor, E major) and outputs the best key path. The path may consist of only one key, or several keys. For example, [F F F F F F] or [Am Am Am C C C] (m indicates minor

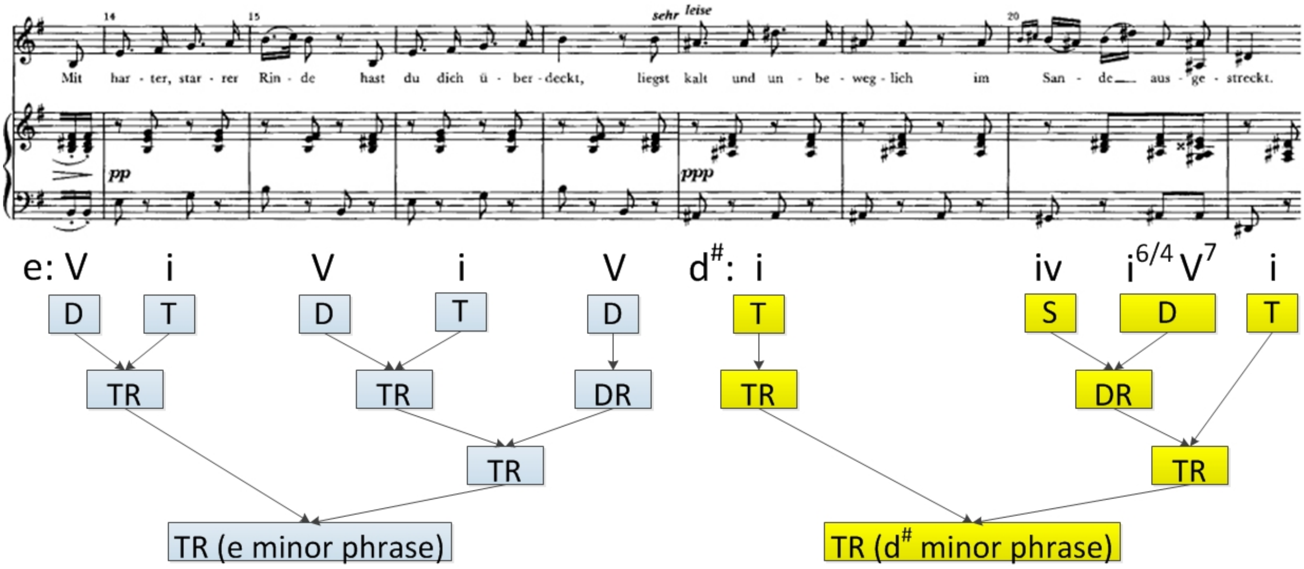

Figure 2. Example of a tonal grammar tree with modulation

chords, other chords are major) are both valid key paths. The *Tonal Grammar Tree* mentioned in §2.1 is the main tool used in this system.

3.2 Algorithm for key partitioning

Each key section is assumed to have at least one tonic chord. The top of each grammar tree must be TR (tonic region), so the key section is a complete tonal grammar tree by itself. Furthermore, the minimum length of each key section is assumed to be 3 chords. However, if no valid paths can be found, key sections with only 2 chords are also considered.

The algorithm is as follows:

1. In a chord sequence, hypothesize any of the chord label as the tonic of a key. Derive the tonal grammar tree of each key.

2. Find if there is any key that can build a single complete tree for the entire sequence. If yes, limit the valid paths to these single-key paths and go to step 7. This phrase is assumed to have a single key only. Else go to next step.

3. For each chord label in the sequence, find the maximum possible accumulated chord sequence length of each key section (up to that label). Determine if this sequence is breakable at that label (The secondary dominant chord is dependent on the subsequent chord. For example, the tonicization segment V/V V cannot be broken in the middle, as V/V is dependent on V chord).

4. Find out all possible key sections with at least 3 chords including at least one tonic chord.

5. Find out all valid paths traversing all the possible key sections, from beginning to end, in a brute-force manner.

Path no.	Key paths					
1	Gm	Gm	Gm	Am	Am	Am
2	Gm	Gm	Gm	C	C	C
3	B♭	B♭	B♭	Am	Am	Am
4	B♭	B♭	B♭	C	C	C

Table 2. All valid key paths in the example

6. If no valid paths can be found, go back to step 4 and change the requirement to "at least 2 chords". Else proceed to step 7.

7. Evaluate the path score of all valid paths and select the one with the highest score to be the best key path.

A simple example is used to illustrate this process. The input chord sequence is [B♭ F Gm Am G C]. Incomplete trees with the keys (B♭, F, Gm, Am, G, C) are built. As all the trees are incomplete, proceed to step 3 and the accumulated length is calculated. The B♭ major tree is shown in Figure 4 as an example. Other five trees (F, Gm, Am, G, C) were built in the same fashion. Either key sections 1-3 or 1-4 of B♭major are valid key sections as they can all be grouped into a single TR and they have at least 3 chords. Then all the valid key paths were found and they are listed in Table 2. All the path scores were evaluated by the equation (1) of the next section.

3.3 Formulation

We have several criteria for choosing the best key path. A good choice of a key section should be rich in tonic and dominant chords, as they are the most important chords to define and establish a key [10]. It is more preferable if the key section starts and ends with the tonic chord, and with less tonicizations as a simpler explanation is better than a complicated one. In a music excerpt, less modulations and modulations to closer keys are preferred. We formulate

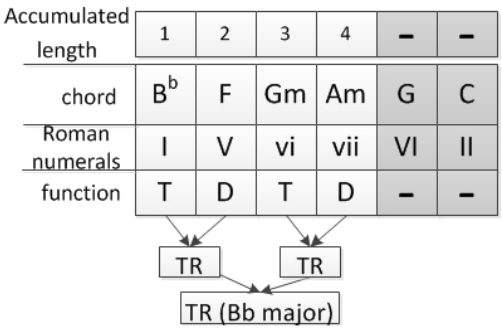

Figure 4. The incomplete B♭ major Tree

these criteria with equation (1):

$$S_{total} = aS_{td} - bS_{ton} - cS_{cost} + dS_{stend} - eS_{sect} \quad (1)$$

where S_{td} is the no. of tonic and dominant chords, S_{ton} is the total number of tonicization steps. For example, in chord progression V/V/ii V/ii ii, the first chord has two steps, while the second chord has one step. $S_{ton} = 2+1+0 = 3$. S_{cost} is the total modulation cost: the total tonal distance of each modulation measured by Lerdahl's distance defined in [8]. S_{stend} indicates whether the excerpt starts and ends with tonic or not. S_{sect} is the total number of key sections. If a key section has only 2 chords, it is counted as 3 in S_{sect} as a penalty. These parameters control how well chords fit in a key section against how often the modulation occurs. S_{td}, S_{ton} and S_{stend} maximizes fitness of the chord sequence to a key section. S_{cost} and S_{sect} induce penalty whenever modulation occurs. The parameters S_{td}, S_{ton}, S_{cost}, S_{stend} and S_{sect} are normalized so that their mean and standard deviation are 0 and 1 respectively. All the coefficients, namely a, b, c, d, e, are determined experimentally, although a slightly different set of values does not have a large effect on the key partitioning results. They are set at $[a, b, c, d, e] = [1, 0.4, 2, 2, 0.4]$. Key structure is generally thought to be hierarchical. An excerpt may have one level of large-scale key changes and another level of tonicizations [17], and the boundary is not well-defined. So it seemed fair to adjust these parameters in order to match the level of key changes labeled by the ground truth. The key path with the highest S_{total} is chosen as the best path.

4. EXPERIMENTS

4.1 Settings

To test the system, we have chosen the Kostka-Payne corpus, which contains classical music excerpts in a theory book [5]. This selection has 46 excerpts, covering compositions of many famous composers. They serve as representative examples of classical music in common practice period (around 1650-1900). All of the excerpts were examined. This corpus has ground truth key information labeled by David Temperley [1]. The mode (major or minor) of the key was labeled by an experienced musician. The chord labels are also available from the website, with the mode added by the experienced musician [2]. All the chord types have been mapped to their roots: major or minor. There are 25 excerpts with a single key and 21 excerpts with key changes (one to four key changes). The longest excerpt has 47 chords whereas the shortest excerpt has 8 chords. The instrumentation ranges from solo piano to orchestral. As we assume the input chord sequence to be harmonically complete, the last chord of excerpts 9, 14 and 15 were truncated as they are the starting chord of another phrase. There are 866 chords in total. For every excerpt, the partitioning algorithm in §3.2 is used to obtain the best path.

4.2 Baseline system

To the best of author's knowledge, there is currently no key partitioning algorithm directly use chord labels as input. To compare the performance of our key partitioning system, another system based on Krumhansl's harmonic-hierarchy information and dynamic programming were set up. Krumhansl's key profile has been used in many note-based key tracking systems such as [3, 9]. Here Krumhansl's harmonic-hierarchy ratings (listed in Chapter 7 of [6]) are used to obtain the perceptual closeness of a chord in a particular key. A higher rating corresponds to a higher tendency to be part of the key. As a fair comparison, the number of chords in a key section is restricted to be at least three, which is the same in our system. To prevent fluctuations of the key, a penalty term $D(x, y)$ is imposed on key changes. The multiplicative constant of penalty term α is determined experimentally to give the best result. The best key path is found iteratively by the dynamic programming technique presented by equations (2) and (3):

$$A_x[1] = H_x[1] \quad \forall x \in K \quad (2)$$

$$A_x[n] = \max \begin{Bmatrix} A_x[n-1] + H_x[n], \\ A_y[n-1] + H_x[n] - \alpha D(x, y) \end{Bmatrix} \\ \forall x, y \in K, \; where \; y \neq x \quad (3)$$

$H_x[n]$ is the harmonic-hierarchy rating of the n^{th} chord with the key x. $A_x[n]$ is the accumulated key strength of the n^{th} chord when the current key is x. K is the set of all possible keys. $D(x, y)$ is the distance between keys x, y based on the key distance in [6] derived from multidimensional scaling. The best path can be found by obtaining the largest A_x of the last chord and tracking all the way back to $A_x[1]$. The same Kostka-Payne corpus chord labels were used to test this baseline system. The best result was obtained by setting $\alpha = 4.5$.

4.3 Results

The key partitioning result of our proposed system and the baseline system were compared against the ground truth provided by Temperley. Four kinds of result metrics were used. The average matching score is shown in Figure 5.

[1] http://www.theory.esm.rochester.edu/temperley/kp-stats/

[2] All the chord and key labels can be found here: https://drive.google.com/file/d/0B0Td6LwTUL-vMVJ6MFcyYWsxVzQ/edit?usp=sharing

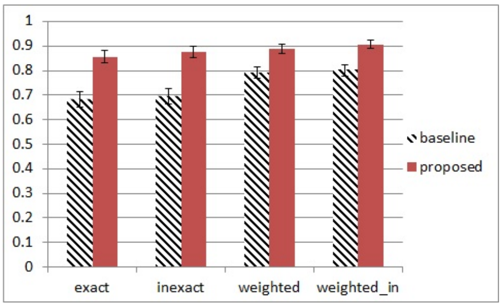

Figure 5. Key partitioning result, with 95% confidence interval

Exact indicates the exact matches between the obtained key path and the ground truth. As modulation is a gradual process, the exact location of key changes may not be definitive. It is more meaningful to consider *Inexact*. For *inexact*, the obtained key is also considered as correct if it matches the key of the previous or next chord. *MIREX* refers to the MIREX 2014 Audio key detection evaluation standard [3]. Harmonically close keys will be given a partial point. Perfect fifth is awarded with 0.5 points, relative minor/ major 0.3 points, whereas parallel major/ minor 0.2 points. This is useful as sometimes a chord progression may be explainable by two different related keys. *MIREX_in* refers to the MIREX standard, but with the addition that the points of previous or next chord will also be considered and the maximum point will be chosen as the matching score of that chord.

The proposed system outperforms the baseline system by about 18% for exact or inexact matching and 0.1 points for MIREX-related scores. It shows that our knowledge-based tonal grammar tree system is better than the baseline system which is based on perceptual closeness. Tonal structural information is exploited, so we have a better understanding of the chord progression and modulations.

4.4 Error analysis

The ground truth key information are compared against the key labels generated by the proposed algorithm. 17 boundary errors were detected, ie. the key label of the previous or next chord was recognized instead. In classical music, modulation is usually not a sudden event. It occurs gradually through several *pivot chords* (chords common to both keys) [10]. Therefore it is sometimes subjective to determine the boundary between two key sections. It may not be a wrong labeling if the boundary is different from the ground truth. Other types of error are listed in Table 3.

The most common error is the misclassification as dominant key, which is the closest related key [10]. It shares many common chords with the tonic key. From Table 4, the same chord sequence can be analyzed by two keys that are dominantly-related. Although the B♭ major analysis contains more tonicizations, the resultant score disadvantage may be outweighed by the cost of key changes, if it is followed by a B♭ major section.

[3] http://www.music-ir.org/mirex/wiki/2014:Audio_Key_Detection

Key relation	*Semitone difference*	*total no.*	*%*
Dominant	7	35	32.7
Supertonic	2	32	29.9
Relative	3	11	10.3
Parallel	0	11	10.3
Minor 3^{rd}	3	9	8.4
Major 3^{rd}	4	8	7.5
Leading tone	1	3	2.8
Tritone	6	2	1.9

Table 3. Eight categories of the 107 error labels

chord symbols	Gm	C	F	B♭	Gm	C	F
F major	ii	V	I	IV	ii	V	I
B♭ major	vi	V/V	V	I	vi	V/V	V

Table 4. Analysis with two different keys

Modulations between keys that are supertonically-related (differs by 2 semitones) or relative major / minor have a similar problem as the dominant key modulation. Many common chords are shared among both keys, so it is easy to confuse these two keys. It is worth to mention that nine of the supertonically-related errors came from excerpt 45. In Temperley's key labels, the whole excerpt is labeled as C major with measures 10-12 considered as a passage of tonicization. However, in [5], it was written that *"Measures 10-12 can be analyzed in terms of secondary functions or as a* ***modulation****"*. If the measures 10-12 are considered as a modulation to D minor, then the analysis of these nine chords is correct.

The parallel key modulation, for example from C major to C minor, has a different problem. Sometimes composers tend to start the phrase with a new mode (major or minor) without much preparation, as the tonic is the same. Fluctuation between major and minor of the same key has always been common [10]. When the phrase information is absent, the exact position of modulation cannot be found by the proposed system.

In another way, there may exist some ornament notes that obscure the real identity of a chord, so that the chord symbol analyzed acoustically is different from the chord symbol analyzed structurally or grammatically. For example, in Figure 6, the first two bars should be analyzed as IV^6-$vii^{\phi 7}$-I progression in A major. However, the C♯ of the I chord is delayed to the next chord. The appoggiatura B♯ made the I chord sound as a i chord, the tonic minor chord instead. Similarly, the last two bars should be analyzed as $IV^{6/5}$-vii^{o7}-i in F♯ minor. However, the passing note A♯ made the i chord sound as a I chord, the original A is delayed to the next chord. In these two cases, the key derived by the last chord in the progression is in conflict with the other chords. Hence the key will be recognized wrongly if the acoustic chord symbol is provided instead of the structural chord symbol.

5. DIFFICULTIES

The biggest problem of this research is lack of labeled data. To the best of our knowledge, large chord label database

Figure 6. Excerpt from Mozart's Piano Concerto no. 23, 2^{nd} movement

for classical music is absent. The largest database we could find is the Kostka-Payne corpus used in this paper. In the future, we may consider manually label more music pieces to check if the system works generally well in classical music.

Moreover, key partitioning is sometimes subjective to listener's perception. In some cases, there are several pivot chords to establish the new key center. "Ground truth" boundaries of key sections are sometimes set arbitrarily. Or there are several sets of acceptable and sensible partitions of key sections. This problem is yet to be studied. Inconsistency between acoustic and structural chord symbols mentioned in §4.4 is also yet to be solved. For any rule-based systems, exceptions may occur. Composers may deliberately break some traditions in the creative process. It is not possible to handle all these exceptional cases.

6. FUTURE WORK AND CONCLUSION

We have only considered major and minor chords in this paper. As dominant 7^{th} and diminished chords are common in classical music, we may consider expanding the chord type selection to make chord labels more accurate. The current system assumes chord labels to be free of errors. We plan to study the method of key tracking in the presence of chord label errors. Then we may incorporate this system to the chord classification system for audio key detection, as the key and chord progression is interdependent. Currently the input phrases must be complete in order to make this tree building process work. We plan to find the key partition method for incomplete input phrases. A more efficient algorithm for tree building process, instead of brute-force, is yet to be discovered. Then less trees are required to be built.

In this paper, we have discussed the uses of tonal grammar to partition key sections of classical music. The proposed system outperforms the baseline system which uses dynamic programming on Krumhansl's harmonic-hierarchy ratings. This tonal grammar is useful for tonal classical music information retrieval and hopefully more uses can be found.

7. REFERENCES

[1] A. Cadwallader and D. Gagné. *Analysis of Tonal Music: A Schenkerian Approach*. Oxford University Press, Oxford, 1998.

[2] B. Catteau, J. Martens, and M. Leman. A probabilistic framework for audio-based tonal key and chord recognition. *Advances in Data Analysis*, (2005):1–8, 2007.

[3] E. Gómez and P. Herrera. Estimating The Tonality Of Polyphonic Audio Files: Cognitive Versus Machine Learning Modelling Strategies. In *ISMIR*, pages 1–4, 2004.

[4] B. Hyer. Key (i). In S. Sadie, editor, *The New Grove Dictionary of Music and Musicians*. Macmillan Publishers, London, 1980.

[5] S. M. Kostka and D. Payne. *Workbook for tonal harmony, with an introduction to twentieth-century music*. McGraw-Hill, New York, 3rd ed. edition, 1995.

[6] C. L. Krumhansl. *Cognitive Foundations of Musical Pitch*. Oxford University Press, New York, 1990.

[7] K. Lee and M. Slaney. Acoustic Chord Transcription and Key Extraction From Audio Using Key-Dependent HMMs Trained on Synthesized Audio. In Array, editor, *Ieee Transactions On Audio Speech And Language Processing*, volume 16, pages 291–301. Ieee, 2008.

[8] F. Lerdahl. *Tonal pitch space*. Oxford University Press, Oxford, 2001.

[9] H. Papadopoulos and G. Peeters. 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, May 2012.

[10] W. Piston. *Harmony*. W. W. Norton, New York, rev. ed. edition, 1948.

[11] C. Raphael and J. Stoddard. Functional harmonic analysis using probabilistic models. *Computer Music Journal*, pages 45–52, 2004.

[12] M. Reger. *Modulation*. Dover Publications, Mineola, N.Y., dover ed. edition, 2007.

[13] M. Rohrmeier. Towards a generative syntax of tonal harmony. *Journal of Mathematics and Music*, 5(1):35–53, Mar. 2011.

[14] H. Schenker. *Free Composition*. Longman, New York, London, 1979.

[15] A. Shenoy and R. Mohapatra. Key determination of acoustic musical signals. *2004 IEEE International Conference on Multimedia and Expo (ICME) (IEEE Cat. No.04TH8763)*, pages 1771–1774, 2004.

[16] R. Steblin. *A history of key characteristics in the eighteenth and early nineteenth centuries*. University of Rochester Press, Rochester, NY, 2nd edition, 2002.

[17] D. Temperley. *The cognition of basic musical structures*. MIT Press, Cambridge, Mass., 2001.

[18] D. H. Younger. Recognition and parsing of context-free languages in time n^3. *Information and Control*, 10(2):189–208, 1967.