Key Estimation Using Circle of Fifths

Takahito Inoshita and Jiro Katto

Graduate School of Fundamental Science and Engineering, Waseda University,
3-4-1 Ohkubo, Shinjuku-ku, Tokyo, 169-8555 Japan
{inoshita,katto}@katto.comm.waseda.ac.jp

Abstract. This paper presents a novel key estimation method of sound sources based on the music theory known as "circle of fifths". We firstly overview music theory and formulate the musical key analysis by vector operations. In detail, we separate music sources into small pieces and calculate FFT-based chroma vectors. They are converted to tonality vectors and COF (circle-of-fifth) vectors are calculated from the tonality vectors, which are mapped onto the circle of fifths coordinate. As a result, each music source can be represented by traces of COF vectors, which usually stay inside a single key region on the circle of fifths. Finally, HMM is applied to the traces of COF vectors in order to detect keys and their boundaries. Experiments using music databases are also carried out.

1 Introduction

Key Estimation is one of important methods in the field of automatic music transcription system. We can estimate the music key by applying macroscopic analysis of the harmony of a tune. On the other way, we can also estimate a chord progression and more microscopic features of a tune. These features may be also useful for music retrieval system.

Music keys are determined from the pitch set in a sufficient span of a tune. So far, many machine learning methods had been proposed [1][2][3]. Most of them apply HMM (Hidden Markov Model) to specific features of the sound such as MFCC (Mel-Frequency Cepstrum Coefficient) and chroma vectors for key estimation or chord estimation, in which EM algorithm is utilized to set up the HMM model by using labeled data. However, since the keys are independent of sorts of music instruments and melody lines, such supervised learning methods work well only for limited sorts of instruments. Furthermore, to make learning data is troublesome and the efficiency depends on the way to learn the data.

In this paper, we propose a new method for extracting features of tunes based on musical knowledge known as “Circle of Fifths” (COF), which projects a tune onto COF coordinate. We firstly separate music sources into small pieces and calculate FFT-based chroma vectors. They are converted to tonality vectors and COF vectors are calculated from the tonality vectors. Finally, they are mapped onto the circle of fifths coordinate (COF coordinate). As a result, each music source can be represented by traces of COF vectors. We then apply HMM to the traces of COF vectors to track music keys and detect their boundaries.

B. Huet et al. (Eds.): MMM 2009, LNCS 5371, pp. 287–297, 2009.

In this research, we deal with only major keys. Minor keys have three scales (natural minor, harmonic minor and melodic minor) and a scale of the natural minor key consists of the same pitch set as its parallel key (A minor's parallel key is C major). Therefore, in this paper, when we talk about a key, we mean both of the key itself and its parallel key.

2 Proposal

Keys of a tune are determined by the pitch set in a sufficient span (usually several bars) of the tune. Generally, they are determined by which notes are used among twelve notes (C, C#, D, …, B) in one octave. For instance, in a C major tune, seven notes [C, D, E, F, G, A, B] are mainly used. Our proposed method is to estimate their pitch sets efficiently. In Fig.1, the process overview of our proposal is shown.

cut out sound pieces from music source

apply FFT and generate chroma vectors

convert chroma vectors into tonality vectors

project tonality vectors onto COF coordinate (COF vectors)

track their plots and estimate keys and their boundaries

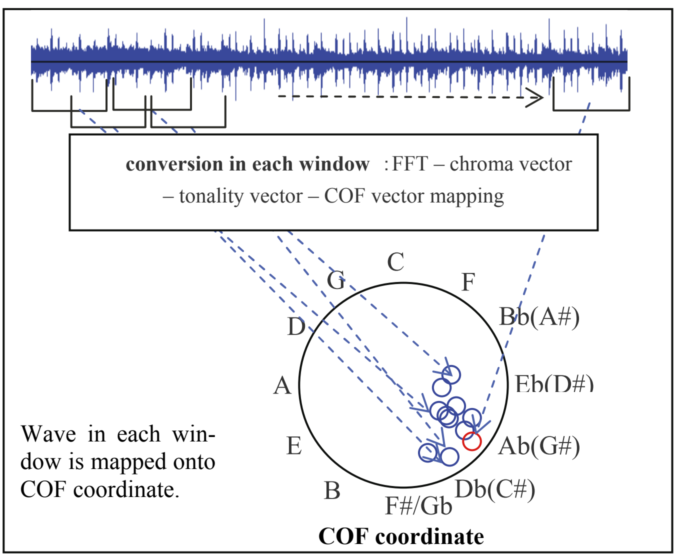

Fig. 1. Flow chart and mapping onto COF coordinate of the proposed method

2.1 Chroma Vector

A chroma vector is calculated from the frequency spectrum and represents how strong power each pitch has as defined in next equations.

$$\mathbf{C}(t) = \begin{bmatrix} c_C(t) \\ \vdots \\ c_B(t) \end{bmatrix}, \ |\mathbf{C}(t)| = 1, \tag{1}$$

$$c_K(t) = \int_{-\infty}^{\infty} BPF_K(\Psi(x))dx \tag{2}$$

where $\Psi(x)$ is the frequency spectrum of a tune and BPF_K is a filter which passes only frequency corresponding of the pitch $K(\in \{C, C\#, ..., B\})$ through. In the equation (1), we normalize a chroma vector because we want to prevent unfairness such that the vector in a span of large volume predominantly works in following processes. Therefore, a chroma vector shows the ratio of power in twelve notes in a span.

2.2 Tonality Vector

A tonality vector is calculated from a chroma vector and represents a probability of each key. So, it is a twelve dimension vector given by

$$\mathbf{K}(t) = \begin{bmatrix} k_C(t) \\ \vdots \\ k_B(t) \end{bmatrix}, \tag{3}$$

$$k_{P_n}(t) = \sum_{i=0}^{11} w_i\, f(c_{P_{i+n(\mathrm{mod}\,12)}}(t)) \quad (P_0 = C, \cdots, P_{11} = B) \tag{4}$$

where w_i is a weight to calculate a tonality vector from chroma and the tonality vector reflects weighted gravities of twelve notes in a certain key. In Spiral Array Model [4], the relation between keys and pitches is defined. In this research, we assign the weight according to the parameters of Spiral Array Model. Concretely speaking, the relation between key *K* and note *n* is measured through the primary triad *T* that connects them (A triad is musical harmony that consists of three notes. Primary triads are three important triads in a key. For instance, C triad, F triad and G triad are primary triads in C major.) As primary triad *T* corresponds to specific key *K* and note *n* does to specific primary triad *T*, the relation between *n* and *T* becomes higher. If there are more than one candidate of *T* for *n* and *K*, all the relation between *n* and *T* are summed up. For instance, the key G major and the note D have very high relation because D is the 5th note of the tonic and the root note of the dominant in the key. On the other hand, the key F# major and the note D have little relation because D is never the member note of any primary triad in the key. So, the weight of D in G major is bigger but, by contraries, the one in F# major is smaller.

In the equation (4), $f(\bullet)$ is a trimming function given by

$$f(x) = \frac{1}{6\left(1 + e^{24(1/12 - x)}\right)}. \quad (5)$$

Its shape is shown in Fig.2.

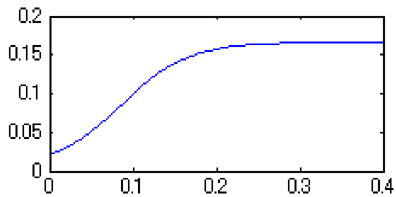

Fig. 2. Trimming function for tonality vector

By using this function, we can restrain unnecessary deviance of a tonality vector by a sudden peak of one note. For example, a case that only the note E appears in C major is supposed. E is an important member note in C major. However, if the chroma vector has a big value only on the element E, the tonality vector has a different feature from C major. In order to prevent this, a note whose power is over a threshold is restrained so that the power doesn't contribute to unexpected keys.

2.3 COF Vector

COF is one of music knowledge and expresses the relation of twelve keys. Twelve keys are put on circumference like Fig. 3, where the neighboring two keys have a similarity that six notes among seven notes of the pitch set is commonly used and only one note differs in semitone. For Example, C major is very similar to F major because C major has [C, D, E, F, G, A, B] as its member notes and F major has [C, D, E, F, G, A, Bb] as its member notes. On the other hand, C major is very dissimilar to F# major (whose member notes are [C#, D#, E#, F#, G#, A#, B]), therefore the two keys face each other.

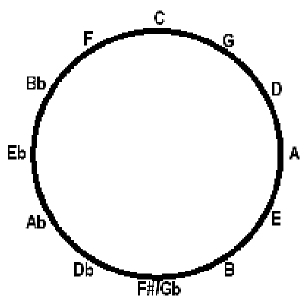

Fig. 3. Circle of Fifths

A COF vector is calculated from a tonality vector and mapped onto COF coordinate. A COF vector is a two-dimensional vector to express tonality in a certain span with similarity of keys. It is given by

$$COF(t) = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = uK(t) \tag{6}$$

where u is a set of twelve unit vectors that represent direction of all the keys. It is given by

$$u = \begin{bmatrix} \cos\left(0\times\frac{\pi}{12}+\frac{\pi}{2}\right), & \cdots & ,\cos\left(11\times\frac{\pi}{12}+\frac{\pi}{2}\right) \\ \sin\left(0\times\frac{\pi}{12}+\frac{\pi}{2}\right), & \cdots & ,\sin\left(11\times\frac{\pi}{12}+\frac{\pi}{2}\right) \end{bmatrix}. \tag{7}$$

How to calculate COF vectors is visibly explained as follows. First, we assume that COF has a two-dimensional coordinate and define unit vectors for the direction of each key (Fig.4.(a)). Next, each unit vector is multiplied by the corresponding element in the tonality vector (Fig.4.(b)). Finally, a COF vector is given as a center of gravity vector of them (the bold arrow in Fig.4.(b)).

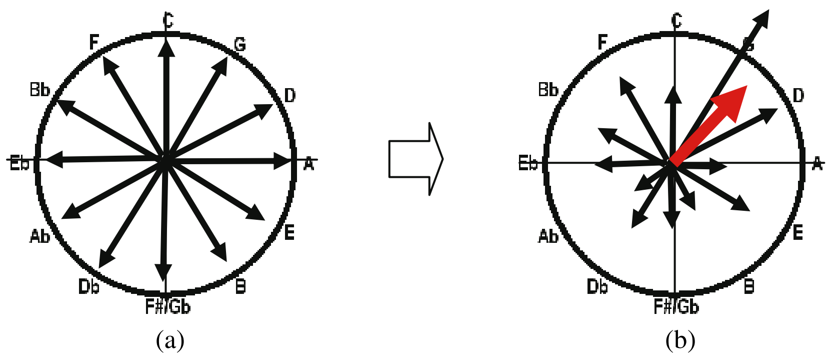

Fig. 4. (a) Unit vectors for the direction of each key. (b). COF vector which is the center of gravity vectors.

We can map one twelve-dimensional tonality vector to one vector on a plane. We call this two-dimensional vector "COF vector". This COF vector presents a key at a certain span by the direction of the vector. In the example of Fig.4.(b), the COF vector shows that the present key may be D major or G major. The COF vector also presents density of the key by the length of the vector. For example, if some long vectors pointing to similar direction exist, then the length of the COF vector becomes long. On the contrary, if elements of a tonality vector are scattered, the COF vector points to neighborhood of the origin. Harmony such as diminish chords or augmented chords corresponds to the latter case.

2.4 Judgment of Key Boundaries

We can get Fig.5 by projection of tonality vectors onto COF coordinate. Plots in the figure correspond to tips of COF vectors. We call these plots "COF plots".

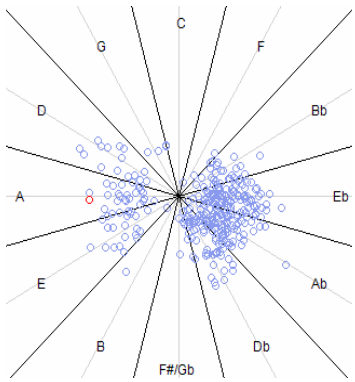

Fig. 5. Series of COF Plots (on COF Coordinate)

In Fig.6, COF plots along time axis are shown. The horizontal axis denotes time and the vertical one does angles of COF vectors on COF coordinate (standard (zero) angle is the direction of C major and positive direction of angle is counterclockwise).

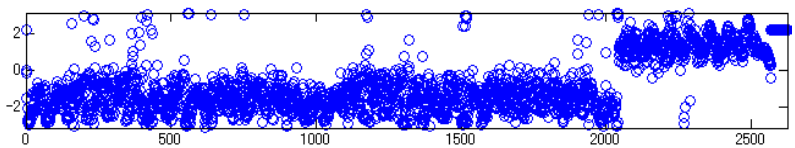

Fig. 6. Series of COF Plots (along Time Axis)

We aim to judge the key boundaries and to identify the keys for this series of COF plots. We want to identify stable regions of swinging plots and judge the translation promptly. For this purpose, we use Hidden Markov Model as a method to estimate keys and their boundaries. We expect that, different from MFCC or chroma vectors, COF plots are easy to handle, robust to track and independent of music instruments when we apply HMM.

2.4.1 HMM

HMM is one of probability models. It is a method to detect unknown (hidden) parameters from observable information. HMM is usually used in the field of speech recognition, genomics and also in music analysis [1][2][3]. HMM is suitable to detect patterns of sequential and flexible signals.

2.4.2 Parameter Settings for HMM

To estimate keys of a tune in our method, the following items are HMM model parameters.

(a) State Set

Each state corresponds to each key of C major, C# major, …, or B major. There are twelve states, for which we don't define a particular initial state and a final state.

$$\begin{aligned} S_i &= S(key_i) \qquad (i = 0,1,\cdots,11) \\ key_i &\in \{Cmajor, C\#major, \cdots, Bmajor\} \end{aligned} \tag{8}$$

(b) State Transition Probability

State transition is equal to key translation, namely modulation. We formulate this effect so that the close two keys on COF coordinate are easy to modulate to each other and the distant two are hard. This rule is defined by next equation (9),

$$a_{ij} = P(S_i S_j) = \frac{C}{e^{p\theta_{ij}} + e^{-p\theta_{ij}}} \tag{9}$$

where a_{ij} is a state transition probability from S_i to S_j, and θ_{ij} is an angle between key_i and key_j on COF coordinate. p is a penalty parameter about modulation, which means that the modulation hardly happens for larger p. C is a constant parameter to normalize the sum of a_{ij} for all possible transitions to 1. The graph of state transition function is shown in Fig.7. The horizontal axis means an angle difference between the current state (key) and the former state (key) and the vertical axis means the probability of the state transition.

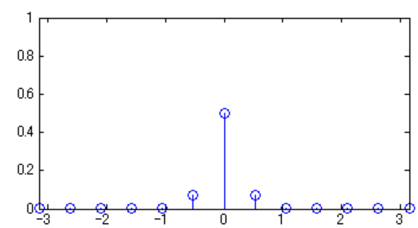

Fig. 7. State Transition Function

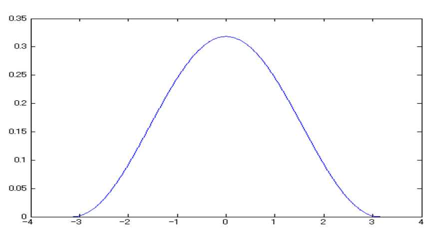

Fig. 8. Output Probability

(c) Output Probability

$f(o;S_i)$ is a probability that symbol o is output in state S_i. In our formulation, symbol o is a plot in COF plots. For all $S_i (i = 0,1,\cdots,11)$, it is necessary for $f(o;S_i)$ to be calculated. In this paper, $f(o;S_i)$ is defined according to how often COF plot o appears in state S_i ,

$$f(o;S_i) = \frac{1}{2\pi}(1+\cos(\theta_{diff})) \tag{10}$$

where θ_{diff} is given by

$$\theta_{diff} = \arg(S_i) - \arg(o)\,. \tag{11}$$

In the equation (11), $\arg(\bullet)$ is an angle that represents direction and independent of its magnitude. The output probability is expressed as shown in Fig.8, where the horizontal axis is θ_{diff} and the vertical axis is corresponding $f(o;S_i)$. It is assumed that, when a plot is distant from the direction of a key, appearance probability of the plot is low.

(d) Output Signal Sequence

Defined by next equation. It corresponds to COF plot series.

$$\mathbf{O} = \{o(t)\} \qquad (0 \le t \le T) \tag{12}$$

(e) State Sequence

Defined by next equation. It corresponds to the key series.

$$\mathbf{S} = \{S(t)\} \qquad (0 \le t \le T) \tag{13}$$

Estimating keys of a tune is equal to detecting the state sequence $\mathbf{S}$ from the output signal sequence $\mathbf{O}$. We use Viterbi algorithm to calculate the output signal sequence and to determine the most probable path.

Note that it is possible to learn the above parameters from training sets by using EM algorithm similar to conventional HMM approaches. However, we apply deterministic (and heuristic) equations as above because the purpose of this paper is to evaluate basic performance of our approach. Parameter learning by using actual data sets and performance comparison are further study.

3 Experiments

We implemented our proposed method by using C# and MATLAB, and executed experiments on a personal computer. We used thirty popular music pieces from the RWC Genre Database [5] and the RWC Popular Music Database [6], and ten classic music pieces from the RWC Classic Music Database [6]. In Fig.9, a result of projection onto COF coordinate is shown. We can see that there are clear clusters of COF plots in the directions of keys of the tune.

Furthermore, graphs of COF plot series along time axis are shown in Figs.10 and 11. We can see the harmony transition along time. Integers on the vertical axis represent the number of sharps of the key signature (flats are counted as negative numbers). For example, C major: 0, E major: 4, and Bb major: 2. Overlaid lines represent keys estimated by HMM.

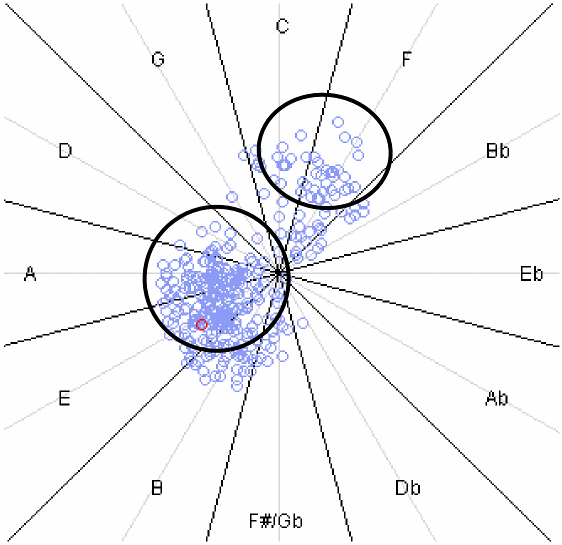

Fig. 9. Projection onto COF Coordinate in case of No.18 song from the RWC Popular Music Database that modulates from E major to F major

The ratio of correct answers is approximately 70%. Popular music pieces and Classic music pieces have similar results. The correct answers are given by trained people's listening to the tunes. In the following chart 1, correct keys of some music pieces we used and their ratio of correct answers are shown.

No. (in RWC Genre Database)	Genre	Correct Key(s)	Ratio of Correct Answers
1	Popular	G	87.40%
2	Popular	C	65.20%
4	Popular	[A-C-E]×2-Em-E	72.10%
6	Popular	Ab-A	97.60%
58-1	Classic	F#	62.60%
58-2	Classic	A	100.00%
59	Classic	D-A-G-D-Dm-D	60.30%
60	Classic	G	82.10%
61	Classic	Bb-Db-Bbm-Bb	52.30%
62	Classic	F#	100.00%
63	Classic	Db-Gb-Db-Db-E-Db	93.00%

Chart 1. Correct keys of some pieces and ratio of correct answers

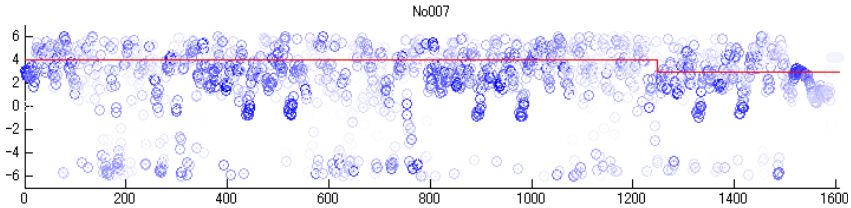

Fig. 10. Key Estimation by HMM (No.7 from the RWC Popular Music Database)

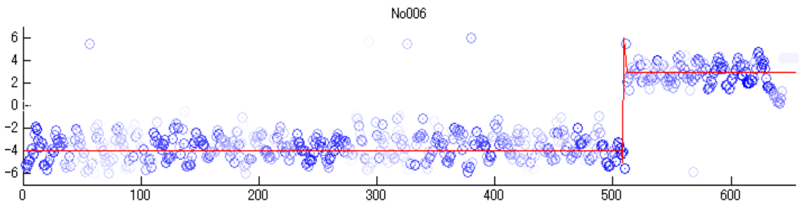

Fig. 11. Key Estimation by HMM (No.7 from the RWC Music Genre Database)

There is little difference of correct answer rate between popular music and classical music. The correct answer rate of music pieces that have frequent modulation or minor keys is lower. The reason No.2 has low correct answer rate is that the singer sometimes sings with blue note scale. Notes of blue note scale is different from notes of a diatonic (ordinary) scale, so it is thought that the result of No.2 is not good.

4 Conclusion

In this paper, we presented a new key estimation method which is independent of music genre and sorts of music instruments. The efficiency was shown by experiments for actual music sources. In our research, we didn't discriminate major keys and natural minor keys, but if we introduce minor keys of other two scales, more robust estimation will be expected. Furthermore, when we use more detailed COF vectors to estimate keys, we can regard them as microscopic features of a tune and apply them to music retrieval system.

References

[1] Lee, K., Slaney, M.: Automatic Chord Recognition from Audio Using an HMM with SupervisedLearning. In: ISMIR 2006 (2006)

[2] Noland, K., Sandler, M.: Key Estimation Using a Hidden Markov Model. In: ISMIR 2006

[3] Cabral, G., Pachet, F., Briot, J.-P.: Automatic X Traditional Descriptor Extraction: The Case of Chord Recognition. In: ISMIR 2005 (2005)

[4] Chew, E.: The Spiral Array: An Algorithm for Determining Key Boundaries. In: Anagnostopoulou, C., Ferrand, M., Smaill, A. (eds.) ICMAI 2002. LNCS, vol. 2445, p. 18. Springer, Heidelberg (2002)
[5] Goto, M., Hashiguchi, H., Ni-shimura, T., Oka, R.: RWC Music Database: Music Genre Database and Musical Instrument Sound Database. 2002-MUS-45-4 2002(40), 19–26 (2002)
[6] Goto, M., Hashiguchi, H., Ni-shimura, T., Oka, R.: RWC Music Database: Popular, Classical, and Jazz Music Databases. In: Pro-ceedings of the 3rd International Conference on Music Information Retrieval (ISMIR 2002), October 2002, pp. 287–288 (2002)