CHORD RECOGNITION USING DOUBLY NESTED CIRCLE OF FIFTHS

Aiko Uemura and Jiro Katto

Department of Computer Science, Waseda University
3-4-1 Ohkubo, Shinjuku-ku, Tokyo 169-8555, Japan
{uemura, katto}@katto.comm.waseda.ac.jp

ABSTRACT

This paper presents a chord recognition method from music signals using chroma vectors and musical knowledge known as "Doubly Nested Circle of Fifths (DNCOF)". DNCOF represents the relationships of major and minor chords where the neighboring two triads are similar. We obtain a novel feature from chroma vectors by mapping them onto two-dimensional DNCOF coordinate, which we call "DNCOF vectors". We expect that the DNCOF vectors can contribute to correcting false recognition obtained by the chroma vectors when their mapped positions are apart from one another in the DNCOF coordinate. In this research, we evaluated our proposal using the Beatles' datasets and showed its effectiveness.

Index Terms— Music, Acoustic signal analysis, Music theory, Chord recognition, Hidden Markov Model

1. INTRODUCTION

Chords are one of the important elements in Western tonal music. Automatic chord recognition is very useful for analyzing musical contents such as melody, keys and harmony. In addition, chord recognition has been focused on in automatic transcription and music information retrieval because automatic chord recognition can be regarded as a pre-processing of automatic transcription. It is also useful to obtain chord labels and their boundaries in cover song identification because chord sequences of cover songs are often analogous.

In general, the feature used for chord recognition is a 12 bin chroma vector which is collection of spectral powers corresponding to octave. This is proposed by Fujishima, who presented a real-time chord recognition system using the chroma vector from DFT [1]. Since then, there have been some approaches to improve the chroma vector [2-4]. Harte et. al proposed a method for tuning the chroma vector to avoid from distributing its spectral power to other bins [2]. Other approaches tried to combine various musical features to calculate better chroma vectors. Ellis showed a chroma vector which is synchronized with beat [3]. Mauch and Dixon used this beat synchronization and, moreover, they applied Non-Negative Least Square problem to spectrogram [4]. As a result, they obtained clearer chroma vectors.

There are basically three ways of recognizing the chord. First, authors of [1,2] proposed methods to use templates defining a chord pitch power based on the chroma vectors. In their studies, 12-element bit masks were used where 1 represents a chord note existence and 0 does non-existent. Oudre et. al extended this approach and used templates of 6 harmonics of every note of the chord [5]. Second, authors of [6,7] proposed HMM based methods which consider the chord labels as hidden values in the model. Bello and Pickens introduced Sheh's work using music theory [8], where they set initial values of state transition probabilities of HMM based on chord distance. Third, there are hybrid proposals of the above methods that combine training data and music theory [9,10]. Some works were devoted to estimate musical contents such as keys and base lines from audio [11-13]. Each element relates closely to the chord, so it is expected that they can be used as a clue to improve the chord recognition rate.

We ourselves had tried to extract key information from audio signals based on the music theory known as Circle of Fifths [13]. We generated key information by abstracting chroma vectors to be mapped onto Circle of Fifths which means key similarity. We extend this method to chord recognition in this paper. In this study, we obtain a novel feature from chroma vectors known as DNCOF and call them "DNCOF vectors". We expect that DNCOF vectors can correct false recognition estimated by chroma vectors and their combined usage contributes to improving the chord recognition rate.

This paper is organized as follows. Section 2 presents our proposal along with definitions of a chroma vector, a chord vector and a DNCOF vector. Section 3 provides experimental results for Beatles datasets and proves effectiveness of our proposal. Finally, Section 4 concludes this paper and refers to future work.

2. METHODS

In this section, we provide our proposed method. Figure 1 shows an overview of our proposal, in which an HMM is prepared for chord recognition and a DNCOF vector is

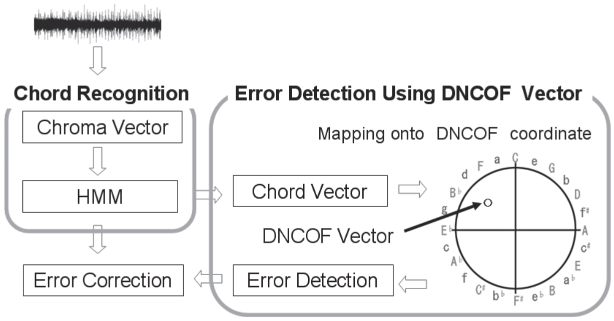

Fig 1. Overview of the proposed method.

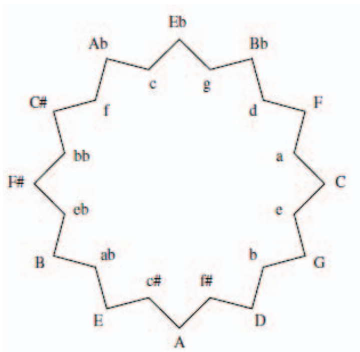

Fig 2. Doubly Nested Circle of Fifths (DNCOF) [8] (The capital letter represents major chord and the lower case represents minor chord)

obtained in three steps from audio signals. First, 12-bin chroma vectors are calculated with beat tracking support. Second, initial chord estimation is carried out and estimation scores for 24 major/minor chords are collected to generate a chord vector. Third, the chord vector is mapped onto DNCOF coordinate and a DNCOF vector is achieved. Finally, the DNCOF vector is utilized for error detection and correction of the recognized chords

2.1. Chroma vector

A chroma vector is collection of spectral powers given by 12 or 36 bins and represents how strong power each pitch has. In our proposal, 12-bin chroma vectors are calculated using ISP (Intelligent Sound Processing) tool box [14]. We align chroma vectors using a spectrum centered on 400Hz. We also apply beat tracking for the chroma vectors to be synchronized with beat similar to [3].

2.2. Chord vector

A chord vector is calculated from the chroma vector as a twenty-four dimensional vector. For this purpose, we prepare a HMM for chord estimation using chroma vectors (its detail is given in 2.4.1). We treat only major and minor chords in this study because most studies [5,7,9,10] consider 24 chords. Accordingly, the HMM is trained for these chords only and produces 24 chord estimation scores (i.e. probabilities) for the chroma vector. A chord vector is defined as a collection of these probabilities and is given by

$$C(t) = \begin{bmatrix} C_C(t) \\ \vdots \\ C_{B_{\min}}(t) \end{bmatrix} \tag{1}$$

$$C_{P_n}(t) = p(P_n \mid chroma(t))$$

$$P_1 = C, P_2 = C\#, \cdots, P_{24} = B_{\min}$$

2.3. DNCOF vector

DNCOF [8] shown in Figure 2 is one of music theories and represents the relationships among major and minor chords where the neighboring two triads are similar.

A DNCOF vector is calculated from a chord vector by mapping it onto DNCOF coordinate. A DNCOF vector is a two-dimensional vector, and expresses a chord component based on similarity. It is calculated as:

$$DNCOF(t) = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = UC_{DNCOF}(t) \tag{2}$$

$$U = \begin{bmatrix} \cos\left(\frac{\pi}{2} - 0 \times \frac{\pi}{24}\right), & \cdots, & \cos\left(\frac{\pi}{2} - 23 \times \frac{\pi}{24}\right) \\ \sin\left(\frac{\pi}{2} - 0 \times \frac{\pi}{24}\right), & \cdots, & \sin\left(\frac{\pi}{2} - 23 \times \frac{\pi}{24}\right) \end{bmatrix}$$

$$C_{DNCOF}(t) = [C_C(t) \quad C_{E\min}(t) \quad \cdots \quad C_{A\min}(t)]^{\mathrm{T}}$$

where U is a matrix determined by a set of twenty four unit vectors, and $C_{DNCOF}(t)$ is obtained by swapping vector components of $C(t)$ according to chord location on the DNCOF in Figure 2.

A DNCOF vector can be also represented in polar coordinates by:

$$\begin{bmatrix} r(t) \\ \theta(t) \end{bmatrix} = \begin{bmatrix} \|[x(t), y(t)]\| \\ angle([x(t).y(t)]) \end{bmatrix}, \; -\pi < angle(v) \le \pi \tag{3}$$

where $r(t)$ is a magnitude of the vector and $\theta(t)$ is an angle from the y-axis (direction to Cmajor). We can regard $r(t)$ as likelihood of chords and $\theta(t)$ as chord types.

2.4. Error Correction

By observing temporal transition, we can see DNCOF plots as shown in Figure 4. The horizontal axis denotes time and the vertical one does angles of the DNCOF vectors on DNCOF coordinate (standard (zero) angle is the direction of C major and positive direction of the angle is counterclockwise). We expect that DNCOF plots are easy to handle and contribute to detect false results of recognition using chroma vectors.

2.4.1. Chord recognition using chroma vectors

We carry out chord recognition using chroma vectors only firstly. This is accomplished by an HMM using a single Gaussian model for its output probability. In our model, we use two chord types, major and minor chords. We use maximal gamma values (which are chord likelihood when chroma vectors are observed) from the forward-backward algorithm instead of Viterbi algorithm to decide chord class.

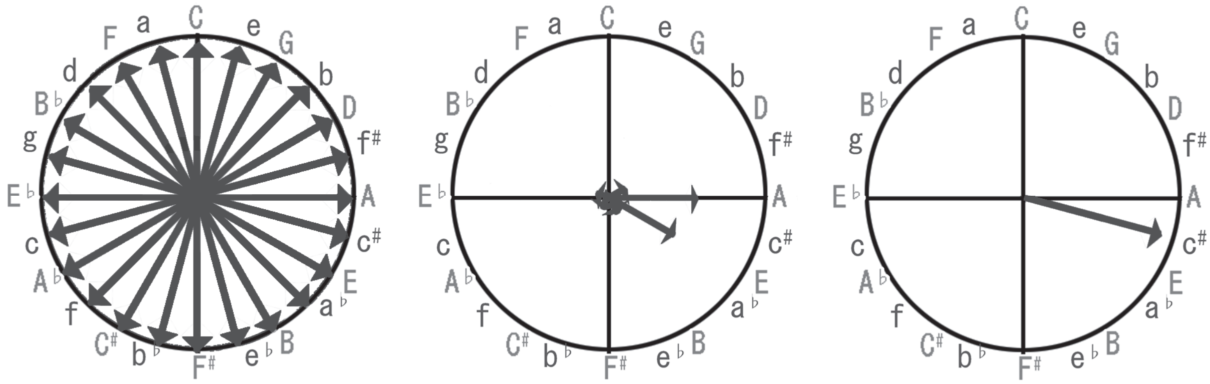

Fig 3. How to map onto DNCOF coordinate (left: a set of twenty four unit vectors, center: multiplying a chord vector by left figure, right: obtain a DNCOF vector)

2.4.2. Error correction using DNCOF vectors

We focus on the distance between plots by the chroma vectors and those by the DNCOF vectors in DNCOF order. This is because their positions are apart from one another when the results are false. In fact, the HMM based on chroma vectors have some correct candidates in its 2nd or 3rd scores even when the results are false. This can be exploited for error correction if we can handle chord vectors and DNCOF vectors in an adequate manner. Otherwise, the distance of chroma and DNCOF vectors remains large.

We treat labels of the chroma vector's results as chord types (angle) and maximal gamma values of the HMM as likelihood (magnitude) in DNCOF coordinate. The distances are calculated by cosine distance frame-by-frame. The threshold processing is used for the error judgment. The threshold of distance is 0.966 in this experiment.

In this paper, error correction is carried out using simple compensation methods. We compare candidates of the chroma-based HMM and the angle of the DNCOF vector in error frame. We choose the nearest label from chroma candidates in DNCOF order and conclude that the most probable candidate in this frame. In addition, we use former chroma values to correct errors when errors occur in chord temporarily.

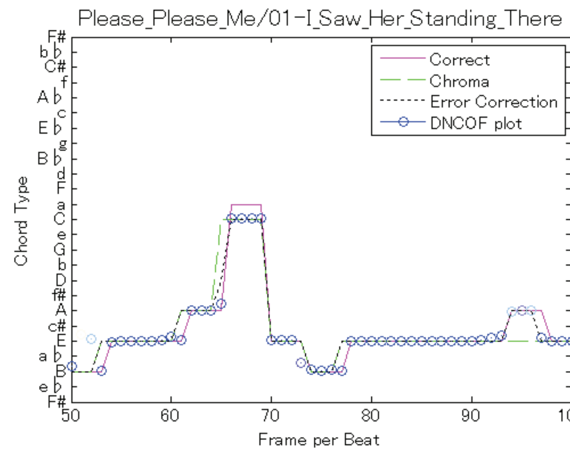

Fig 4. Each result from the Beatles' I Saw Her Standing There (from 50 to 100 frames per beat)

Table 1. Results of each compensation method

compensation method	Accuracy
none	66.7%
DNCOF only	67.5%
combined	**68.1%**

3. EVALUATIONS

We implement our proposed method using MATLAB and examine experiments on a personal computer. Audio signals are downsampled to 16 kHz and separated into small pieces. The frame size consists of 4096 samples. We used annotated Beatles dataset as described in isophonics [15], which consists of 180 Beatles' songs. It provides labels other than major and minor chords. We group triads and other chords with the same root into the same category. For instance, we treat C minor triad and C minor augment chord as C minor chord.

3.1. Error Detection

To evaluate effectiveness of our error detection methods, we computed error detection rates of 180 Beatles songs. The detection rate is **71.5%** in this system. This shows that DNCOF vectors can determine whether errors have occurred in chord recognition using chroma vectors. Figure 4 supports this result which shows how far apart the DNCOF vectors and chroma results are. When the errors have occurred, we can see that each DNCOF plot is slightly displaced from chroma one as shown in Figure.4. In addition, the threshold value is 0.966. This represents that the difference of angle corresponds to one in DNCOF order (one angle = $\pi/12$) when errors have occurred.

3.2. Chord Recognition

Though trained data and test date are prepared separately in many recognition experiments, we follow MIREX audio chord estimation style [16] in this paper. We computed recognition rates using 3-fold cross validation.

We evaluate the next three methods described in subsection 2.4.2.

(1) Chord recognition using chroma (no compensation)
(2) Compensation using DNCOF vector
(3) Compensation using DNCOF and former value

Table1 shows the recognition-rate accuracy in percentage. The combined case often shows the best performance among compared methods. The errors tend to have intervals and occur in chord boundaries. The errors are slightly improved at the chord boundaries when using combined compensation. The compensation values are apart from correct ones if transition happens from one chord to distant chord in DNCOF order.

Figure 4 supports this result which shows how far apart the DNCOF vectors and chroma results are. When the errors have occurred, we can see that each DNCOF plot is slightly displaced from chroma one as shown in Figure.4. We can recognize that the chroma outputs are not enough and the correct answer remains in other. Therefore, the chroma based HMM can be improved with the help of DNCOF if we can correctly adjust the false output.

On the other hand, some plots are not corrected in Figure 4. We can see that the error correct path is drawn by chroma path. This may be due to the mapping method for DNCOF vectors abstracted from chroma vectors, and the fact that DNCOF vectors depend on chroma vectors.

4. CONCLUSIONS

In this paper, we proposed an approach for chord recognition using DNCOF vectors and chroma vectors. Our results showed that the method has efficiency for improving the recognition rate and DNCOF vectors can detect false recognition frames. This indicates that the DNCOF vectors have efficient features for audio analysis.

However, the DNCOF vectors leave much to be improved. In order to accomplish this goal, we enhance the precision of DNCOF vectors and re-evaluate the method to improve chroma vectors themselves. Furthermore, we should obtain error intervals clearly and examine the better methods for error correction.

The output of the DNCOF vectors will be useful in automatic transcription and cover song identification. Furthermore, we will propose an application for understanding and training several chords because DNCOF vectors enable to visualize chord similarity.

5. REFERENCES

[1] T. Fujishima: “Real-time chord recognition of musical sound: A system using common lisp music,” Proc. ICMC, pp. 464-467, Oct.1999.

[2] C. Harte and M. Sandler: “Automatic chord identification using a quantised chromagram,” in Proc. Audio Eng. Soc., Spain, May.2005.

[3] D. P. W. Ellis and G. Poliner: “Identifying cover songs with chroma features and dynamic programming beat tracking“, Proc. ICASSP, pages IV 1429-1432, 2007.

[4] M. Mauch and S. Dixon: "Approximate Note Transcription for the Improved Identification of Difficult Chords," Proc.ISMIR, Aug.2010.

[5] Oudre. et.al: “Template-Based Chord Recognition : Influence of the Chord Types,” Proc. ISMIR, pp. 153-158, Oct.2009.

[6] A. Sheh and D. P. Ellis: “Chord segmentation and recognition using EM-trained hidden markov models,” Proc. ISMIR, pp. 183-189, Oct.2003.

[7] K.Lee and M.Slaney: “Acoustic chord transcription and key extraction from audio using key-dependent HMMs trained on synthesized audio,” IEEE Trans.on Audio,Speech and Language Processing,16(2), pp.291-301, 2008.

[8] J. P. Bello and J. Pickens: “A robust mid-level representation for harmonic content in music signal,” Proc. ISMIR, pp. 304‼ 311, Sep.2005.

[9] M.P.RyynHanen and A. P. Klapri: “Automatic transcription of melody, bassline, and chords in polyfonic music,” Computer Music Journal, 32(3), pp.72-86, 2008.

[10] H. Papadpoulos and G.Peeters: “Large-scale study of chord estimation algorithms based on chroma representation and HMM,” Proc. the International Workshop on Content-based Multimedia Indexing, pp53-60, June.2007.

[11] T. Yoshioka. et al, “Automatic Chord Recognition Based on Probabilistic Integration of Chord Transition and Base Pitch Estimation”, Proc. ISMIR, Oct.2008.

[12] T.Rochear et al.: “Concurrent Estimation of Chords and Keys from Audio," Proc.ISMIR, pp141-146, Aug.2010.

[13] T. Inoshita and J. Katto: “Key Estimation using Circle of Fifths”, 15th International Multimedia Modeling Conference, Jan.2009.

[14] Intelligent Sound Processing : http://kom.aau.dk/project/isound/

[15] isophonics : http://isophonics.net/

[16] MIREX (Music Information Retrieval Evaluation eXchange): http://www.music-ir.org/mirex/wiki/MIREX_HOME