Research on CRFs in Music Chord Recognition Algorithm

Feng Wang
College of Information Engineering, Taiyuan University of Technology Taiyuan, China
Email: wangfeng@tyut.edu.cn

Xueying Zhang
College of Information Engineering, Taiyuan University of Technology Taiyuan, China
Email: tyzhangxy@163.com

***Abstract*—In this paper, five different methods in music recognition are discussed, a new character MPCP is proposed in Chord Recognition. The new character overcome the limitation of the traditional PCP and MFCC, apply for recognition system by combining both characteristics. For features we use MPCP vectors, it is trained by CRFs, finally we acquired accurate chords by Viterbi. The experiment show that the proposed strategy can reach a good performance in chord recognition and proved our strategy to be quite promising.**

***Index Terms*—chord recognition, Pitch class Profile (PCP), Mel PCP, Mel Frequency Cepstrum Coefficient (MFCC), Conditional Random Fieids (CRFs)**

I. Introduction

Chord is an important part of the music signals, which is compose of three or more tones and appear simultaneously or near simultaneously[1]. Chords fully represent the content and features of a music, it is the image of the most important means of modeling the music. so chord is the soul of the music. Therefore, identification of chords for the whole understanding of music is very important, nowadays chord recognition methods based statistical have been applied broadly and acquired some impacts. Caozheng[2] presented arithmetic based on template mating and decision trees with max likelihood coefficient, having better recognition rate , although the algorithm is relatively simple. But he only recognized single chord, which have no practicality. Recent studies have shown the benefit of applying stochastic modeling to this task ,they will create a label text noted the cord boundaries and type, so you can easily identify. And then through a random model to training and identifying files, so to achieve full recognition of the song[3][4] [5] [6].

In chord recognition methods based statistic, there have two keys need solved: one is segmentation of chord sequences, the other is that choosing and recognizing the features. Any period of musical composition by a number of chords, different chords in music played different roles, the performance of their ways and the effect of the music is very important. How to distinguish the different chords precisely, is not only the basis for understanding the whole song, but also to avoid the mistaken. Feature selection is particularly important for the identification of chords, chord feature is not only has good recognition performance, but also have the corresponding support for music theory.

This paper aims at the characteristics of chords and their role in the music, bring forward a chord recognition method based MPCP(Mel Pitch class Profile) feature: this feature founded music theory, simultaneity divided up by Mel frequency, having better auditory characteristics. The songs recognized are the label files made by Chris Harte, this label file is a text file format, chord boundaries and types are be noted which performed by the Beatles[7].

II. MPCP (Mel Pitch class Profile)

A. PCP *Feature*

Position figures and tables at the tops and bottoms of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be centered below the figures; table captions should be centered above. Avoid placing figures and tables before their first mention in the text. Use the abbreviation “Fig. 1,” even at the beginning of a sentence.

A chromagram or a Pitch Class Profile (PCP) is the choice of the feature set in automatic chord recognition or key extraction since introduced by Fujishima[8]. Perception of musical pitch involves two dimensions – height and chroma[9]. Pitch heightmoves vertically in octaves telling which octave a note belongs to. On the other hand, chroma tells where it stands in relation to others within an octave. A chromagram or a pitch class profile is a 12-dimensional vector representation of a chroma, which represents the relative intensity in each of twelve semitones in a chromatic scale. Since a chord is composed of a set of tones, and its label is only determined by the position of those tones in a chroma, regardless of their heights, chroma vectors appear to be an ideal feature to represent a musical chord or a musical key.

doi:10.4304/jcp.8.4.1016-1019

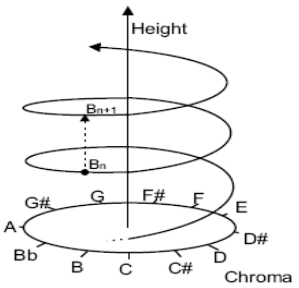

Figure 2. The Pitch Class Profile.

The audio files are downsampled to 11025Hz, divided into overlapping frames of N = 4096 points and converted to a short-time Fourier transform (STFT) representation.

$$X_{STFT}(k,n) = \sum_{m=0}^{N-1} x(n-m) \cdot w(m) \cdot e^{-j2\pi km/N} \tag{1}$$

We use a finer grained PCP vector of 12 dimensions to give some flexibility in accounting for slight variations in tuning. A step size of 100ms, or 10 PCP frames per second, is employed. STFT bins k are mapped to PCP bins p according to:

$$p(k) = [12 \cdot \log_2(k/N \cdot f_{sr}/f_{ref})] \bmod 12 \tag{2}$$

Where f_{ref} is the reference frequency corresponding to PCP[0] and f_{sr} is the sampling rate. For each time slice, we calculate the value of each PCP element by summing the magnitude of all frequency bins that correspond to a particular pitch class i.e.

$$PCP(p) = \sum_{k:p(k)=p} |X(K)|^2 \quad p = 0,1\cdots 11 \tag{3}$$

B. MPCP *Feature*

As chord feature, PCP represents music theory, but could not consider factor that auditor is recognizer. Having following limitation: 1: In the low-frequency characteristics of rather vague, impacting recognition; 2: computing complexity great, affecting the recognition efficiency; 3: easy to confusion in the peak; 4: In the fast-rhythm, easy to empty chord under the mistaken identification. Some researchers attempt to recognize chore by MFCC feature, Its effect is not good. The reason are: 1: MFCC Only consider the characteristics of the human ear's hearing did not take into account characteristics of music theory; 2: MFCC has a good recognition performance and noise immunity, but it has strict demands to computing and precision;3:MFCC inhibit pitch information, affect the recognition results. Through the above analysis, we can recognise by auditory characteristics and a combination of music theory as the feature, this will be the development direction of chord recognition. Therefore, this paper combines human auditory characteristics and music theory, presenting a new feature: MPCP(Mel PCP).MPCP first make frequency spectrum translate into non-linear spectrum based on Mel frequency spectrum, and then account formula (2)and (3). After Mel frequency-domain transformation MPCP feature Includes not only the theoretical basis of music, but also made full use of the specific perceptual characteristics of the human ear. This characteristic does not depend on the nature of the signal, and does not make any assumptions and limitations to the input signal; Also made use of the auditory model of research results, when the SNR is lowered we still has a good recognition performance.. MPCP calculation shown in Figure 2:

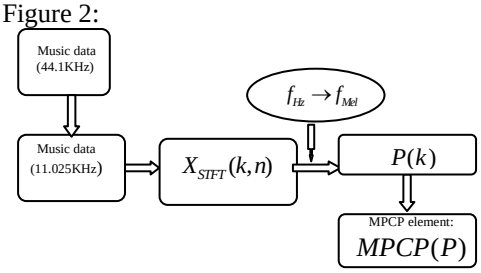

Figure 2. Block Diagram of Computing MPCP Feature.

III. RECOGNITION MODELS

A. *Hidden Markov Models*

HMMs are generative stochastic models that attept to model a hidden first-order Markov process based on a set of observable outputs. Nonetheless, it should be noted that in reality chord progressions are high-order Markov processes, and so the first-order Markov assumption may result in model deficiencies. HMMs for chords recognition have used with two different approaches[4,7].

B. *Conditional Random Fields*

Given a sequence of observations X, HMMs seek to maximize the joint probability $P(X,Y)$ for a hidden state sequence Y. This method works well in practise, but from a theoretical perspective, it is not quite the question that one ought to be asking at recognition time. During recognition, the observation sequence is always fixed, and so it may make more sense to model only the conditional probability distribution $P(Y|X)$. This frees the recognizer from needing to enforce any particular model $P(X)$ of the data, and thus it is no longer necessary for the components of the observation vectors to be conditionally independent, as they must be for HMMs. Such a model may include thousands or even millions of observation features.

The closest analogue to the HMM that uses this modeling technique is known as the linear-chain CRF, one of the most commonly used member of the larger CRF family. Besides the probabilistic characteristics mentioned above, CRFs differ from the HMMs in that each hidden state depends not just on the current observation but on the complete observation sequence. At decoding time, linear-chain CRFs are quite similar to HMMs, using a variant of the Viterbi algorithm. Unlike

an HMM, however, in order to train a linear-chain CRF, one must have access to fully labelled and aligned training data. Training is considerably slower for CRFs than it is for an HMMs regardless of whether one uses a path-discriminant or model-discriminant approach, but fortunately, there are optimisation techniques to improve the training speed. We chose the limited-memory Broyden–Fletcher–Goldfarb–Shanno method, a variant of Newton's method.

IV. Chord Recognition Algorithm Based On MPCP

Because this identification songs are Beatles songs, which belong to the scope of Western music. Although Western Music having a strong sense of Equal Temperaments, However, add-chord rarely appear in their music. In addition to added-chord and diminished-chord, intervals compose of themselves are non-consonance, therefore this two types chords do not belong to non- consonance. Among the musical works, we few use this two chords. At the same time the same root of the chord is also divided into a class, therefore, the final chord sets are divided into large and small two major categories of 24 kinds of chord. For the classification of chord, we recognize chords by 24 states of the HMM. each state being a class of chord.

Algorithm steps are as follows:

First step: Feature Extraction. First we deal with the songs by beat-synchronous tracking proposed by Simon Dixon[10]. We uses training HMM by the MPCP as the feature vector, since each state corresponds to a class of a chord, therefore, the state can distinguish between different chords to achieve the purpose of identifying.

Second step: Training HMM or CRFs. We consider the result of the first received MPCP characteristics and existing tag file together as the HMM input. The label document made by Chris Harte provides all the detailed information of the training data (chords), for the model parameters we can be directly estimate: Through the formula (4-33), (4-34), (4-35) in the reference [11] to estimate the model parameters of each state.

Third step: Chord Recognition. After the previous step of training, each type of chords are generating a kind of model parameters. The Viterbi algorithm is applied to the model to find the optimal path through the given model parameters and input observation vector. In the algorithm implementation, every step of recursive process to save the best condition, it can be recovered upon an optimal state transition sequence after recursion termination, making the greatest probability of observation vector and getting the best recognition results.

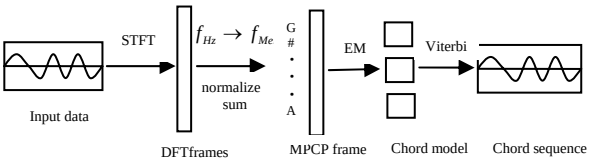

Figure 3. Recognition System.

V. Expermental Results And analysis

Training data are the three Beatles album of 48 songs and Test data are also the three Beatles songs. In this paper, we used PCP and MPCP as the feature respectively, Recognition Results as Table I.

Table I.
Recognition Results

Training data	Testing data	Feature+model	Recognition Rate
Album1: Please Please Me Album 2: Help Album 3: Abbey Road	song 1: Help	PCP+HMM	72%
		MPCP+HMM	80%
		MPCP+CRFs	88.2%
	song 2: I Need You	PCP+HMM	60%
		MPCP+HMM	65%
		MPCP+CRFs	77.5%
	song 3: Yesterday	PCP+HMM	69%
		MPCP+HMM	70%
		MPCP+CRFs	71.7%

As can be seen above table, MPCP characteristic feature is superior to PCP. recognition rate based PCP feature is low there are several reasons for: As the analysis window is fixed, when rhythm speeding up that could jump some empty chords, it caused confusion; The triad chord and seventh chord have the same root note, and seventh chord is a triad on the basis of the root note on the superposition of a seven-degree note, in addition, some seventh and triad contain the same pure-tone, therefore, it easy to error in the recognition process. For example: In the song "Help" appears in 20 or so with the same root note and the same pure tone, which are in PCP feature., similarly, in the remaining two songs of the recognition errors also appears in the above situation. The MPCP features consider the human ear's auditory, And have a better recognition in the changing frequency domain, therefore the recognition rate of MPCP was significantly higher than PCP. It is to some extent overcome the problem of confusion chords and basically solved the wrong recognition of empty chord.

By comparison with recognition rate of the songs 1、2 and 3, we found the latter two songs, in general, the recognition rate of less than 1 song, both PCP and MPCP feature. By comparing the tag files and songs, recognition errors are concentrated in a few chords- Suspension of chord: it is make the three tone of triad higher semitone or whole tone, which is a chord pure 4 degrees that away from the root note formed. Because of this chord change part of interval, for example Suspension of two chord: it used the two degree tone instead of the three degree tone. So it exists unstable tone. The song 2 and 3 contain 28 Suspension of chords, in which there are 26 Suspension of chords in song 2, but song 1 does not appear in the case, so recognition rate of song 1 is higher than the last two songs, and the recognition rate of the song 2 only has 60%.

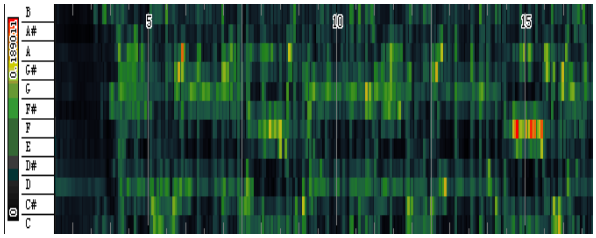

Figure 4. Song Clip.

PCP+HMM is classical chord recognition model, which was proposed initially by Sheh and Ellis. but HMM is memoryless and can not make full use of context information, so the model obtained a local optimal value. CRFs overcome the shortcoming of HMM, it have the expression of long-range dependence characteristics and the ability, that could solve the issues such as labeling (classification) offset, and all features can be globally normalized, global optimal solution can be obtained. However, CRFs have a defect: running time is long.

VI. Summary And Future

This paper presents a chord recognition algorithm based on MPCP feature, that can successfully recognize chords in unstructured. This is a challenging instance of extracting complex musical information from a complex input signal. The paper implements the recognition chords by comparing with feature of PCP and MPCP, verify the new features proposed in this paper has better recognition ability and stability. The next will be based on the above works, and further study impact of musical characteristics in chords; choice and optimization of recognition model.

Acknowledgment

The authors wish to thank A, B, C. This work was supported in part by a grant from XYZ.

References

[1] Ren daming: Course of Music Theory in New Concepts. beijing: People's Music Publishing House,2003

[2] Caozheng. Music Recognition Algorithm Based Time-frequency analysis [D].Hefei:University of Science and Technology of China,2006

[3] Logan B., Salomon A.: A music similarity function based on signal analysis, IEEE International Conference on Multimedia and Expo, Tokio, Japan, 2001

[4] A.Sheh and D. P. Ellis. Chord segmentation and recognition using EM-trained hidden Markov models[J]. In Proceedings of the International Symosium on Music Information Retrieval, Baltimore, MD, 2003.

[5] B. Su, S. Jeng, "Multitimber Chord Classification Using Wavelet Transform andSelforganized Map Neural Networks". In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Salt Lake City, USA, pages 3377 3380, 2001

[6] T. Yoshioka, T. Kitahara, K. Komatani, T. Ogata, H. G. Okuno. Automatic chord transcription with concurrent recognition of chord symbols and boundaries[J]. In Proceedings of Proceedings of ISMIR, 2004

[7] C. A. Harte and M. B. Sandler. Automatic chord identification using a quantised chromagram. In Proceedings of the Audio Engineering Society, Spain, 2005. Audio Engineering Society.

[8] T.Fujishima. Realtime chord recognition of musical sound: A system using Common Lisp Music[C]. In Proceedings of the International Computer Music Conference. Beijing: International Computer Music Association, 1999.

[9] G. H. Wakefield. Mathematical representation of joint time chroma distributions[J]. In SPIE, Denver, Colorado, 1999.

[10] S. Dixon. A lightweight multi-agent musical beat tracking system[J]. In Proceedings of the AAAI Workshop on Artificial Intelligence and Music: Towards Formal Models for Composition, Performance and Analysis. AAAI Press, 2000

[11] Li bingnan Study Of Music Recognition Algorithm Based On RBF Neural Network And HMM [D] Taiyuan Taiyuan University Of Technology, 2009

Feng Wang and the other authors may include biographies and photographs at the end of regular papers. Photographs, if provided, should be cropped into 26mm in width and 32mm in height. The first paragraph may contain a place and/or date of birth (list place, then date). Next, the author's educational background is listed. The degrees should be listed with type of degree in what field, which institution, city, state or country, and year degree was earned. The author's major field of study should be lower-cased.

The second paragraph uses the pronoun of the person (he or she) and not the author's last name. It lists military and work experience, including summer and fellowship jobs. Job titles are capitalized. The current job must have a location; previous positions may be listed without one. Information concerning previous publications may be included. Try not to list more than three books or published articles. The format for listing publishers of a book within the biography is: title of book (city, state: publisher name, year) similar to a reference. Current and previous research interests ends the paragraph.