[[1]](#footnote-1)

A Brief Review of Automatic Chord Recognition

郭锦岳，11510478

# ntroduction

T

he Automatic Chord Recognition, is an important technology in the field of Music Information Retrieval. To explain the function of ACR, we need to first introduce the concept of chord. A chord in musicology, is defined as a set of simultaneous tone, usually more than three tones. If there are three tones in total, the chord is called a triad; if there are four tones, it is called a seventh chord. Due to the definition of musical tones, the multiple frequencies are defined as a same tone. For example, 440 Hz, and 880 Hz, is the same tone *A*. In a typical musical piece, the chord will change frequently, generally changes every few seconds. This changing of chords carries the emotion and progress of a musical piece, hence it is a very fundamental concept in music.

The difficulty of ACR comes from the aspect of acoustics, musical instruments, and musicology.

In the field of acoustics, we found that the harmonic component of instruments in complex. We recognize the pitch of a tone through its fundamental frequency, while retrieve the timbre from other harmonic components. These overtones are unique for different instruments, different playing techniques, and different recording conditions, which makes it difficult to extract the fundamental frequencies.

Since many music pieces are played by multiple instruments it also increases the difficulty. For example, the three tones of a chord might be played by three different instruments; It is common that the lowest tone, called root tone, is played by a bass; a middle tone played by piano, and the highest tone played by violin or trumpet.

The composition of music also increases this difficulty. Although defined as simultaneously played tones, the actual chord might be played separately, and still be recognized as an identical chord. Another problem is that there might be out-of-chord tones in a musical piece. For example, the background piano plays a C major chord, consists of C, E and G; However, a singer might be singing a melody of A, B, D, C.

In this brief review, I will introduce two dominant methods in Automatic Chord Recognition and some new methods. First one is the oldest method called Pitch Class Profile (Fujishima, 1999)[1]. Another enhanced method, which borrows the thought of Automatic Speech Recognition, is applying Hidden Markov Model (HMM), first implemented 2003. (A. Sheh and D. P. Ellis)[2]. Further enhancements, including Enhanced Pitch Profile (EPCP), (K. Lee, 2006)[3], Convolutional Neural Networks (CNN) [4], will be briefly introduced as well.

# Pitch Class Profile

Pitch Class Profile is a specific feature, used in Automatic Chord Recognition. Perception of musical pitch has two dimensions: *height* and *chroma.* Pitch height moves 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. The Pitch Class Profile (also called *chromagram*), is a 12-dimentional vector representation of a chroma, which represents the relative intensity in each of twelve semitons in a chromatic scale. Since a chord is composed of a set of tones, and its label is only determined by the position of these tones in a chroma, regardless of their heights, *chromagram* seems to be an identical feature to represent a musical chord.

In order to obtain the PCP, take procedures as following:

1. For input signal , compute the STFT
2. Compute constant-Q transform , which is a logarithmic compression. It tunes the spectrum to the resolution of human ear. For band , the frequency is , where *B* is manually set, usually 12, 24 or 36.
3. Compute the PCP: *b* is the PCP index from 1 to *B*.

Figure 1 is an example of PCP, also known as chroma vector (generated in [3]):

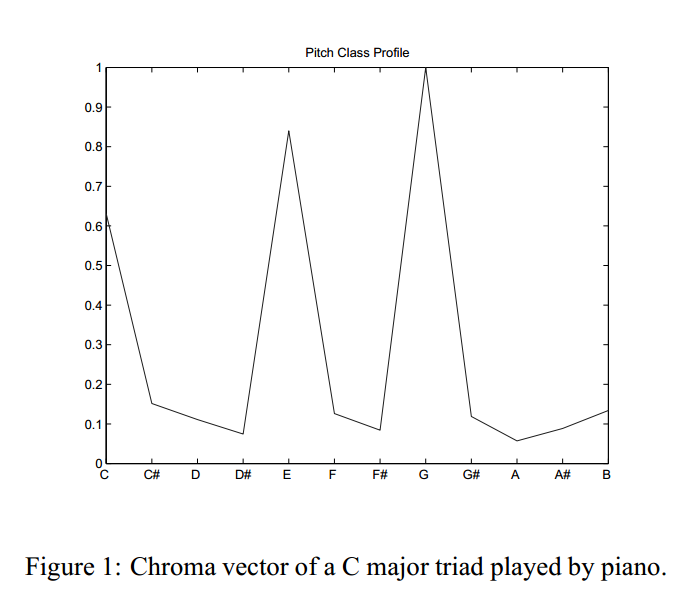


Figure 1

In the original method of [1], the obtained PCP vector will then calculate the correlation with a set of binary chord templates. For example, the template of a C major triad is [1,0,0,0,1,0,0,1,0,0,0,0]. The original simplest template set only contains 24 chords, which are 12 major triads and 12 minor triads.

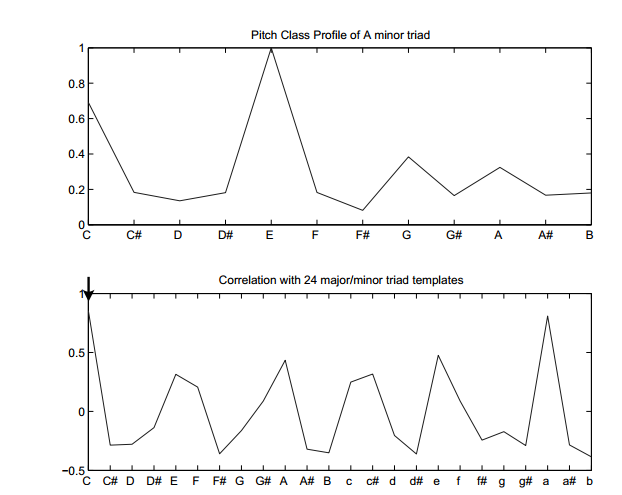


Figure 2: PCP of an A minor triad, and its correlation with the templates. Arrow in the lower figure denotes where the maximum correlation occurs.

Figure 2 is another example in [3] that shows a mistake of this model. After calculating correlations, the PCP of an A minor chord has the greatest correlation with C major template.

The reason of this mistake might be two parts. First we can see the PCP vector at above: the A minor chord should only have the components of A, C and E, but the value of tone G is even greater than A. The reason of this might be those mentioned in the first paragraph. The second reason is that our referencing templates are only binary sets, which is impossible for real input.

Even though, this method of PCP vector and correlation has been the very first in the field of ACR, and has the basic ability to recognize chords in the simplest condition: simultaneous and single-instrument.

# Enhanced Pitch Class Profile

According to the disadvantages of PCP vector, many methods have been developed to enhance the performance. Here I introduce one method called EPCP, which is simple but efficient, developed by (K. Lee, 2006) [3].

As mentioned in the first chapter, instruments’ tone has complex harmonic components, while only the fundamental frequency is useful to us. The traditional PCP method calculates every integer-times frequency, while in the field of music, only 2-power-times harmonic is could be considered as a same chroma.

The author Lee borrows the idea of Harmonic Product Spectrum (Schroder 1968, [5]) to develop a better performance of retrieving fundamental frequency.

The HPS feature and fundamental frequency are generated by equation:

Now, we modulate this equation to fit the definition of musical tones:

Figure 3: DFT of A minor triad example (above) and its Harmonic Product Spectrum (below)

Figure 3 shows the efficiency of HPS in reducing redundant harmonic components. If we use HPS, instead of DFT, to generate PCP vector, the result is defined as Enhanced Pitch Class Profile, EPCP.

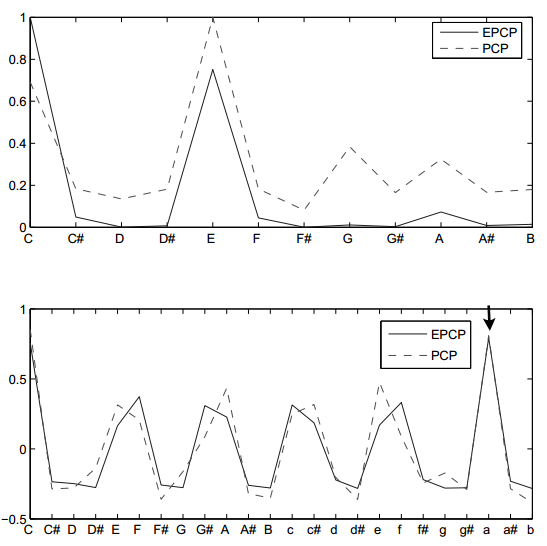


Figure 4: comparison of EPCP and PCP vector (above), and the corresponding correlation result with binary templates (below).

With the same input as Figure 2, we can see the result of EPCP has greatly reduced the mistaken value on G. In the correlation result, we can see the greatest correlation occurs with A minor template, which is the correct result.

However, as mentioned in chapter 1, real musical pieces contains rapid chord changing. Figure 5 is the recognition result of PCP and EPCP of Bach’s *Prelude in C major*:

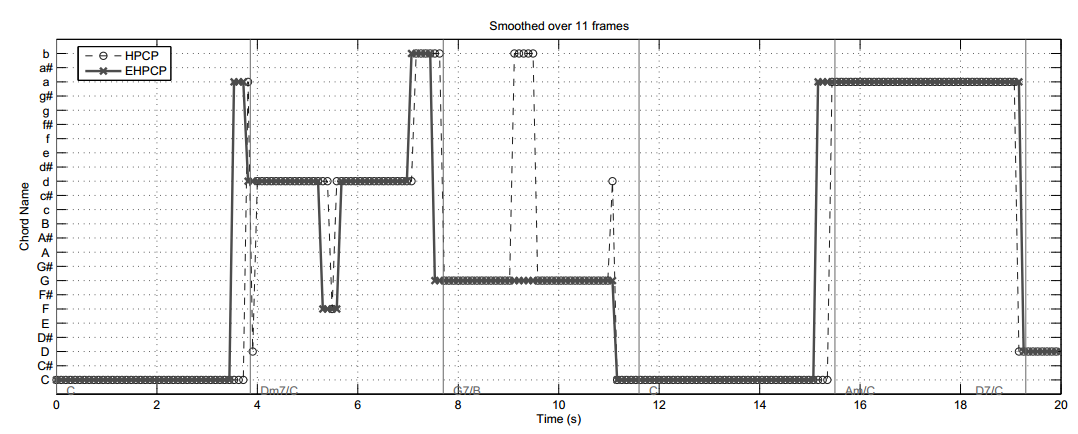


Figure 5: chord recognition results after a smoothing process across 11 frames of an excerpt from Bach’s *Prelude in C major.*

The method of EPCP has reduced many transient errors of traditional PCP method. However it also shows some error, especially before the changing of chord. This is because the composition of Bach. He adds transition tones before chord changes, which makes it difficult to recognize. Moreover, for EPCP, we can observe that most of the errors occur in the second bar. This is because the actual chord in the second bar is Dm7/C, which is a transpose seventh chord. Such a chord does not exist in the 24 templates, so the result changes between several similar chords, but still not satisfying.

# Hidden Markov Model

Previous methods could be classified as Acoustical method, since the information they use is no more than memory-less acoustical information. As mentioned in the first chapter, the chord changing is not a random process. Similar with speech, chord changing has its own context to express certain emotions. For example, a widely used chord progression is C-Am-F-G, that almost more than 50% pop songs contain such a chord progression.

Based on the idea of semantic recognition, scholars again borrow the method of Hidden Markov Model from speech recognition. HMM is a very successful statistical method, which requires a training process.

The basic idea of Markov Model is a discrete state changing chain. For every moment, there is a current state that shows the circumstance of a system. After a certain period of time, the system has a certain probability that turns into another state, or stay in the same state. The probability of state changes in the next moment is based on the current state of this moment.

In the Hidden Markov Model, this state changing could not be directly seen. Instead, we could only guess the state through our observation. We define another set of probabilities, to show the relationship between the real state and our observation. For a certain chain of observation, we could then figure out the most probable state chain.

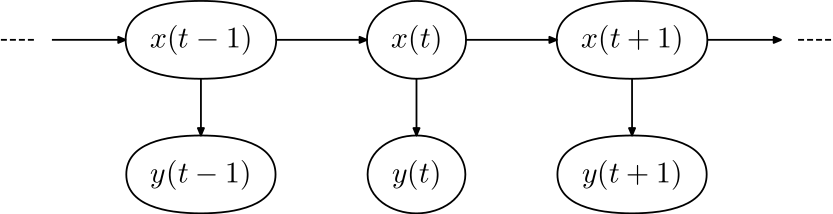


Figure 6: A schematic diagram of Hidden Markov Model. x(t) is the chain of states, while y(t) is the chain of observations.

The HMM method is first introduced to the field of ACR by A. Sheh and D. P. Ellis, 2003[2]. They define the chord labels as the state, and *pitch class profile* as the observation.

Another fundamental idea of HMM is training. We use a training set of data to let the model automatically adjust the two probability sets. After training, we are able to calculate the state chain for arbitrary input by the Viterbi Algorithm.

Here the introduced training process is applied by K. Lee and M.Slaney, 2006 [6].

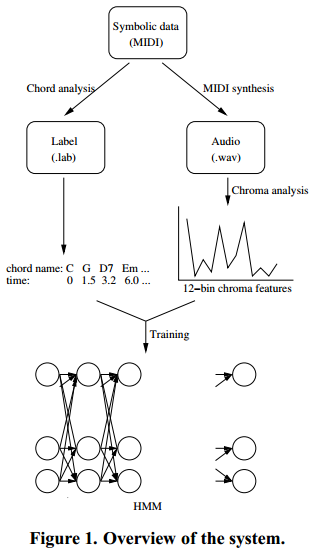


Figure 7: An HMM training system built for ACR.

The input training sets are generated from MIDI files. MIDI is a digital protocol for music signal, it contains digitized information of pitch, duration, timbre, etc. The author first analyzed this digital file to generate the Label set, which is easy if the pitches are known. He then uses the MIDI file to synthesis an audio waveform file, which is done by some Digital Audio Workstation (DAW). The audio file is then used to generate chromagram vectors, which is the pitch class profile in chapter 2.

The labeling data is used as state sets, while the pitch class profile data is used as observation sets. When labeling chords, the author used a 36-dimention matrix, which contains not only major and minor triads, but also diminished triads.

After this training process, two matrix has been established. One is the probabilities of state changing, which is called *transition matrix*; another is the probabilities of an observation with corresponding states, called *covariance matrix.*

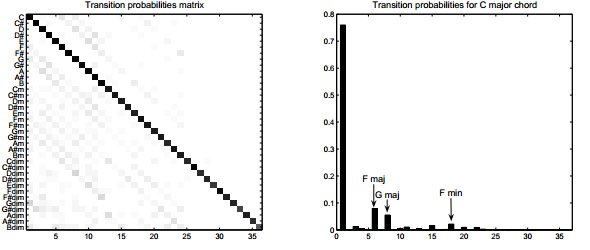


Figure 8: transition matrix (left), and its first row (right)

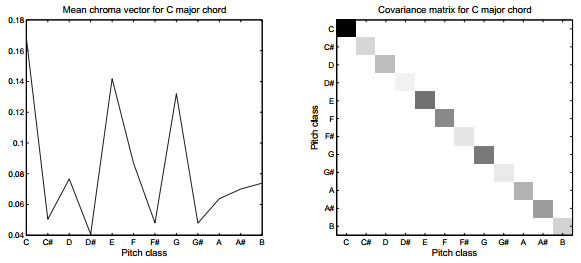


Figure 9: mean chroma vector (left), and covariance matrix (right) of C major chord.

Figure 8 denotes the transition probabilities after training. As in the right figure, we can see that if the current state is C major, the most probable next sate is still C major. The second probable choice is F major, then G major, which is identical with then experience on music theory.

Figure 9 shows the relationship between states and observation. The average observation of C major chord is like the left figure, and the three peaks in the right figure is the three tones in C major, C, E and G.

Furthermore, the author generates a chromagram diagram, which is resemble with spectrogram in speech processing.

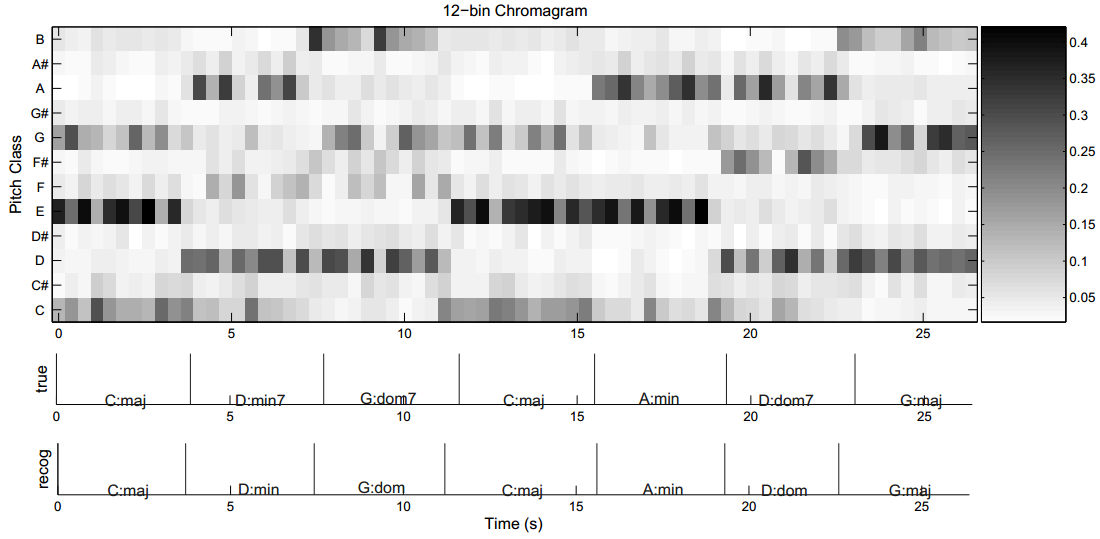


Figure 10: 12-bin chromagram (PCP) with true result and HMM-generated result (below), musical piece from Bach’s *Prelude in C major* performed by Gleen Gould.

Comparing the true value and recognized result, we can see this HMM has a better precision in recognition. Since the template dose not contain seventh chords, the second, third and sixth bars has been degraded to the most similar triad.

# Performances

# Developing methods

# Conclusion

References

1. T. Fujishima, “Realtime chord recognition of musical sound: a system using common lisp music,” In *Proc. ICMC*, 1999
2. A. Sheh and D. P. Ellis, “Chord segmentation and recognition using EM-trained hidden Markov models,” in *Proceedings of the International Symposium on Music Information Retrieval,* Baltimore, MD, 2003.
3. K. Lee, “Automatic chord recognition using enhanced pitch class profile,” in *Proceedings of the International Computer Music Conference*, New Orleans, USA, 2006.
4. E. J. Humphrey and J. P. Bello. Rethinking automatic chord recognition with convolutional neural networks. In *ICMLA 11*, volume 2, pages 357–362, 2012.
5. Schroeder, M. R. (1968). Period histogram and product spectrum: New methods for fundamental-frequency measurement. *Journal of the Acoustical Society of America* 43(4),829–834.
6. K. Lee and M. Slaney, “Automatic chord recognition from audio using an HMM with supervised learning,” in Proc. 7th Int. Conf. Mus. Inf. Ret., 2006, pp. 133–7.

1. [↑](#footnote-ref-1)