A HIERARCHICAL APPROACH FOR MUSIC CHORD MODELING BASED ON THE ANALYSIS OF TONAL CHARACTERISTICS

Namunu C. Maddage Mohan S. Kankanhalli Haizhou Li*
Institute for Infocomm Research, 21, Heng Mui Keng Terrace, Singapore 119613
{maddage, hli}@i2r.a-star.edu.sg
*School of Computing, National University of Singapore, Singapore 117543
mohan@comp.nus.edu.sg

ABSTRACT

This paper first discusses how the signal segmentation and tonal characteristics of music notes effect in music chord detection. Two approaches, pitch class profile approach and psycho-acoustical approach, which differently represent these tonal characteristics, are examined for chord detection. The analysis of the tonal characteristics reveals that not only the fundamental frequency of music note but also its harmonics and sub-harmonies in different octaves contribute for detecting related music chord. A hierarchical approach, which transforms the music chord tonal characteristics in each octave onto probabilistic space, is then proposed for modeling the music chord. Our experimental results show that detection of chord type, Major, Minor, Diminish, and Augmented, and individual chords, 12 chords per chord type, are improved with the proposed hierarchical chord modeling approach. Experimental results also reveal that the tempo proportional signal segmentation is more effective extracting tonal characteristics than using fixed length segmentation.

1. INTRODUCTION

Sequence of music chords describes the harmony line of the music. Detection of harmony line is essential for music structure analysis which is useful for developing music related applications such as music information retrieval systems, music transcription, music streaming, watermarking scheme for music etc. The main steps followed in the existing chord detection systems [1] [6] [9] [11] [13] are described below.

- Signal segmentation where within the segment, the temporal properties of the music chords can fairly be considered stationary (fixed length 20~30ms)
- Feature extraction to characterize the chords – pitch class profile feature (which mainly measures the fundamental frequencies (F0) of the notes that comprise the chord).
- Statistical learning techniques for chord modeling - Hidden Markov Model (HMM), Gaussian Mixture Model (GMM), Support Vector Machine (SVM), Neural Networks (NN).

However, due to polyphonic nature of the music signals, the chord detection has been a challenging problem. In this paper we exam

- The tonal characteristics representation in both pitch class profile (PCP) approach and psycho-acoustic profile (PAP) approach for chord detection problem
- How signal segmentation effects the extraction of tonal characteristics

We then propose a hierarchical approach to model the tonal characteristics to represent music chord. Previous algorithms [1], [9], [11], [13], have been tested on a small set of a few different chords, because it's difficult to find a large database which consists of a large variety of chords. In our experiments, we use synthetically generated music chords (*12 chords of each Major, Minor, Diminish, and Augmented chord type)* in addition to the chords extracted from songs. Thus we can estimate the detection accuracies of many different chords.

There have been many research efforts since early 20th century to find the psychological representation of the pitch. Stevens et al. (1937) [10] described pitch perception as continuous psychological effect, which is proportional to the magnitude of the frequency (i.e. *pitch height*). Goldstein (1973) [2] and Terhardt (1974) [12] proposed two psycho-acoustical approaches: harmonic representation and sub-harmonic representation, for complex tones respectively. In Goldstein's pitch representation, music tone is characterized by fundamental frequency (F0) with harmonic partials. Terhardt suggested that each separable component of a complex tone generates eight sub-harmonics and the frequency of most of the commonly generated sub-harmonics determines the perceived pitch. Laden and Keefe (1989) [5] implemented Goldstein and Terhardt methods for pitch representation and claimed psycho-acoustical representation of music pitches has advantages over pitch class representation in chord type detection (Major, Minor and Diminished) . However they haven't mention statistics to support their claims. Moorer (1975) [7] utilized harmonic information of the tones to identify music notes. Pitch perception experiments conducted by Ritsma (1967) [8] concluded that that the fundamental frequencies in the 100-400Hz range and their 3rd, 4th and 5th harmonics, which cover up to 2kHz frequency range produce well-defined pitch perception in human ears. Ward (1954) [14] acknowledged that the upper limit of the music pitch is in the 4.5 kHz frequency range. Thus, the upper limit of the F0s of tones produced by musical instruments is set below 5 kHz. The highest tone (C7) of the piano has F0 of 4186Hz.

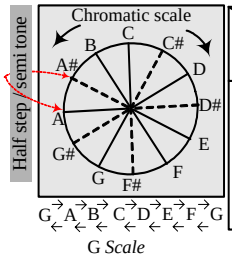
Rest of the paper is organized as follows. Commonly used pitch class profile (PCP) approach and psycho-acoustic profile (PAP) approach for music chord characterization are discussed in section 2. Section 3 details our proposed hierarchical chord modeling approach. Experimental results are discussed in section 4. We conclude the paper in section 5.

2. POLYPHONIC PITCH REPRESENTATION

The feature which characterizes the chord should ideally be insensitive to the source characteristics. In section 2.2 and 2.3 we discuss commonly used pitch class profile approach and psycho-acoustical profile approach for music pitch representation respectively. In section 2.1 we briefly highlight the music knowledge and explain the formulation of music chords.

2.1. Music scale and chord formulation

A set of notes, which forms a particular context and note pitches arranged in ascending or descending order, is called the music scale. The eight basic notes (C, D, E, F, G, A, B, C) which are the white notes in the keyboard, can be arranged in an alphabetical succession of sounds ascending or descending from the starting note. This note arrangement is known as *Diatonic Scale* and it is the most common scale used in traditional western music. Psychological studies have suggested that human cognitive mechanism can effectively differentiate the tones in the diatonic scale (Krumhansl 1979 [4]). In a music scale, the pitch progression from one note to the other is either half step (a semitone -S) or the whole step (a tone –T). Thus, it expands the eight notes into 12 pitch class. The first note in the scale is known as Tonic and it is the keynote (tone-note) from which the scale takes the name. The notes arrangement in G-scale is shown in Figure 1. *Chromatic Scale*, which is cyclic nature in octave periodicities, shares same symbol/value for two tones separated by an integral number of octaves.



Chord type \ Notes	Distance in whole step (T) to the notes from *Key* note		
	1st note	2nd note	3rd note
Major (maj)	0.0T	2.0T	3.5T
Minor (min)	0.0T	1.5T	3.5T
Diminished (dim)	0.0T	1.5T	3.0T
Augmented (aug)	0.0T	2.0T	4.0T

T - Imply a Tone / whole step in music theory

Figure 1: Distance to the notes in the chord from the key note in the scale

Music chords are constructed by selecting notes from the corresponding scales. Types of commonly used chords are Major, Minor, Diminished and Augmented. Each chord type consists of 12 chords. The first note of the chord is the key-note in the scale and Figure 1 shows the note distances to the 2nd and 3rd notes of the chord from the key-note. For example, Gmaj chord has note G, B and D. Table 1 describes the fundamental frequencies (F0s) of the notes in different octaves (C2B2 ~C8B8) based on the ISO standard concert pitch A4=440Hz.

Table 1: F0s of music notes and their positions in the octaves

Octave		~ B1	C2B2	C3B3	C4B4	C5B5	C6B6	C7B7	C8B8	
Freq-range (Hz)			64~128	128~256	256~512	512~1024	1024~2048	2048~4096	4096~8192	
12 Pitch class notes	C	0 ~ 64	65.406	130.813	261.626	523.251	1046.502	2093.004	4186.008	All the higher octaves in the (8192 ~ 22050) Hz frequency range
	C#		69.296	138.591	277.183	554.365	1108.730	2217.460	4434.920	
	D		73.416	146.832	293.665	587.330	1174.659	2349.318	4698.636	
	D#		77.782	155.563	311.127	622.254	1244.508	2489.016	4978.032	
	E		82.407	164.814	329.628	659.255	1318.510	2637.02	5274.04	
	F		87.307	174.614	349.228	698.456	1396.913	2793.826	5587.652	
	F#		92.499	184.997	369.994	739.989	1479.978	2959.956	5919.912	
	G		97.999	195.998	391.995	783.991	1567.982	3135.964	6271.928	
	G#		103.826	207.652	415.305	830.609	1661.219	3322.438	6644.876	
	A		110.000	220.000	440.00	880.000	1760.000	3520.000	7040.000	
	A#		116.541	233.082	466.164	932.328	1864.655	3729.310	7458.62	
	B		123.471	246.942	493.883	987.767	1975.533	3951.066	7902.132	

ISO 16 standard specifies A4 = 440Hz and it is called as concert pitch

2.2. Pitch class profile (PCP) approach

Many of the previous chord detection systems have utilized pitch class approach to represent the music signals [1][3][4][9][11]. In order to construct 12 Pitch Class Profile (PCP) feature vector, first music signal is transformed into frequency domain. Using Eq(1) linear frequency scale (f_{linear}) is mapped into octave scale (f_{octave}) where Fs, N, F_{req} are sampling frequency, number of FFT points and reference mapping point respectively. In our implementation we set frequency resolution (Fs/N) equals to 1Hz, F_{req}=64Hz (F0 of the note C2) and C=12 (12 pitches). Linear to octave frequency mapping depicted in Figure 2.

$$f_{octave} = \left[C * \log_2 \left(\frac{Fs * f_{linear}}{N * F_{ref}} \right) \bmod C \right] \quad \textbf{(1)}$$

Then 12 rectangular filters are placed near notes in each octave to capture the strengths of the F0s of the music notes. PCP vector construction is explained in Eq(2). F0 strengths of the same j^{th} note across all the octaves are summed up to form the j^{th} coefficient of the PCP vector. In the Eq(2), S(.) is the frequency domain magnitude (in dB) signal spectrum. $W_{(OC,j)}$ is the filter whose position and the pass-band frequency range varies with both octave index (OC) and j^{th} note in the octave (OC). If octave index is 1 then the respective octave is C2B2.

$$PCP(j) = \sum_{OC=1}^{8} \left[S(.) W_{(OC,j)} \right]^2 \quad OC = 1....8, \quad j = 1.....12. \quad \textbf{(2)}$$

The reasons for using filters to extract strengths of note F0s, are explained below.

1. Due to physical configuration of the instruments, the F0s of the notes may vary from the standard values (A4=440Hz is used as concert pitch and notes in different octaves are set according to A4).
2. Though the physical octave ratio is 2:1, cognitive experiments have highlighted that this ratio is closed at lower frequencies, but increases with the higher frequencies. It exceeds by 3% at about 2 kHz [14]. Therefore, we position filters to detect the strengths of the harmonics of the shifted notes.

2.3. Psycho-acoustical approach

Earlier research on pitch perception reveals that the central processing unit in the human auditory system responds not only to the F0 of the pitch but also the harmonics and sub-harmonics of the pitch [2][12][14]. F0s of all the music notes in different octaves are described in Table 1. From the table we can see that 3rd and 6th harmonic of C4 is closed to the F0 of G5 and G6 respectively. Similarly, 5th and 7th sub-harmonics of E7 are closed to F0 of C5 and F#4 respectively. Therefore, we place filters around F0s of the notes to capture these sub-harmonic and harmonic strengths. Figure 2 depicts the filter position setting. We position 12 filters in each octave covering 8 octaves (C2B2 ~C8B8). Eq(3) describes the construction of the j^{th} coefficient of the i^{th} Psycho-Acoustic Profile (PAP) feature vector.

$$\begin{aligned} PAP^{\,i}(j) &= \left[S(.) W_{(OC,n)} \right]^2 \quad where \quad j = 1....(12 * OC) \\ OC &= 1 + floor\,(j/12) \\ n &= j - 12 * (OC - 1); \end{aligned} \quad \textbf{(3)}$$

In the orchestra, different musical notes are played in different octaves. As far as PCP feature vector is concerned, it averages

effects of music notes (tonal characteristics) across all the octaves and represents in 12 coefficients. PAP feature vector can be visualized as the expansion of PCP feature vector which considers effects of the notes in all the octaves individually.

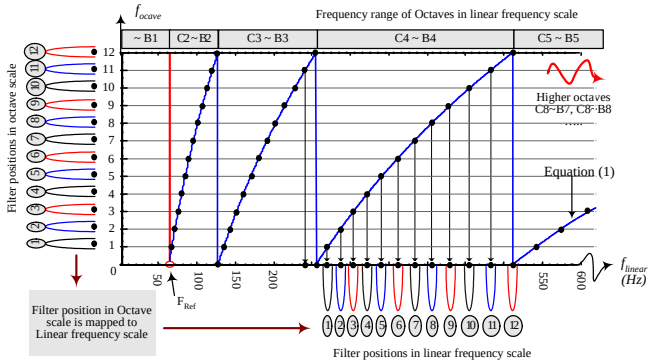


Figure 2: Filter bank distribution in octave scale for tapping harmonics and sub-harmonics of polyphonic pitches.

3. HIERARCHICAL CHORD MODELING

In the state of the art chord modeling, we represent the chord signal's tonal characteristics of all the octaves in a vector for a single layer classifier [1][11][13]. As is noted in our initial experiments depicted in Figure 5 (sec 4), those tonal characteristics in individual octaves alone are capable of detecting chords. We therefore propose 2-layer hierarchical model (see Figure 3). In the first layer, we build individual tonal characteristic model for each octave. The responses of the individual models in the first layer are fed to the model in the second layer to detect the chord.

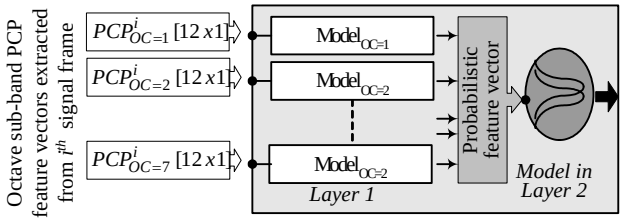


Figure 3: Two layers hierarchical representation of a music chord

The models in the 1st layer are trained using PCP feature vectors (12x1) which are constructed from individual octaves. Note that the tonal characteristics in C9B9 octave (i.e. OC=8 octave index 8) are less reliable, only C2B2~C8B8 octaves are considered. The 2nd layer model is trained with the vector representation of the 1st layer model responses. In our implementation we use 4 Gaussian mixtures for each model in layer 1 and 2. Therefore, the vectors that activate the layer 2 model are probabilistic vectors. This hierarchical chord modeling can be visualized as the transformation of vector space chord modeling (in layer 1) into probabilistic space (layer 2). We then use this 2 layer representation to model 48 music chords.

4. EXPERIMENTS

In our database, the samples (44.1 kHz sampling frequency, 16 bits per sample and mono) of 48 music chords (12 Major Chords, 12 Minor Chords, 12 Diminished chords and 12 Augmented chords) are divided into two clusters.

Cluster 1(CLS1): Chord samples in cluster 1 are generated at the synthetic environment. Since we couldn't collect adequate number of samples for each of 48 chords from real music we generated samples in cluster 1 using Cakewalk software which has rich high quality tone database. These computer generated music chords can be considered as clean samples which are ideal for the ground truth for our experiments. We first generate music notes using Cakewalk software where tempo of the note is varied from 80 to 200 BPM (beats per minutes) in 10 BPM steps, octave variation is C2B2~C7B7 and meter is 4/4. Then we create chords by mixing music notes according to Figure 1. Note mixing procedure for creating a chord is shown in Figure 4.

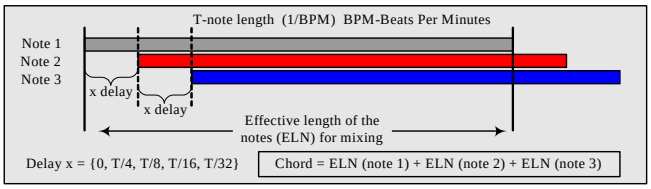


Figure 4: Note mixing procedure for creating a synthetic chord

By setting different delays at note mixing we can generate the same chord differently. As show in Figure 4, we set the delay x to {0, T/4, T/8, T/16, T/32} and mix the notes to generate 5 different samples of each chord. These time delays are set to make the synthetically generate chords as close as to the chords generated in the orchestra. For example, strumming delay of strings to generate a chord is proportional to the tempo of the music. Then we circulate the note mixing such that note 3 will take the position of note 1, note 1 takes the note 2 position and note 2 takes the note 3 position. We totally generated over 2000 samples for each chord.

Cluster 2(CLS-2): Chord samples in cluster 2 are extracted from CD quality 40 English songs (10-Michael Learns To Rock (MLTR), 10-Bryan Adams, 6-Beatles, 8-Westlife and 6-Backstreet Boys). With the aid of music sheets and listening tests have been carried out to annotate the chords songs at 16th note level. Around 70% of chord samples in cluster 2 belongs to major and minor chords. When experiments are conducted on CLS-2 samples we include all the samples in CLS-1 for training the chord models.

We segment the chords into tempo proportional frames and extract both PAP and PCP features. 128 Gaussian mixtures are employed for modeling the chords with cross validation of around 60% of training samples in each turn. Figure 5 shows the average chord detection accuracy when the tonal characteristics are extracted from the individual octaves. The experimental results reveal that tempo proportional segmentation (TPS) of chord signal can improve the average chord detection accuracy. Around 72% and 60% of average accuracy can be achieved in all the individual octaves except octave C8B8 and C9B9, when the tests are conducted on chord samples in CLS-1 and CLS-2 respectively. This test also highlights both sub-harmonics and harmonics effects are useful for chord detection. When we average all the harmonics and sub-harmonic effects across the octaves, then it becomes the PCP representation the chord. When effects are spread across the octaves then it becomes PAP representation of the chords. However, in above experiment PCP feature vector is equal to PAP feature vector because only the effects of 12 notes in the octave are considered

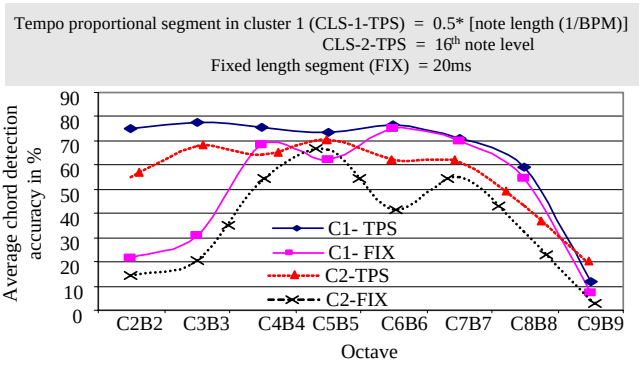


Figure 5: Chord detection accuracies when individual octaves are considered

Based on the above test results we consider C2B2~C8B8 octave range for constructing PCP and PAP features. The average chord detection accuracies are shown in Figure 6. It can be seen that the pitch class representation of chord, where note effects (F0, sub-harmonic, harmonic) are averaged across the octaves performed better than psycho-acoustical representation of chords. Our proposed 2 layer hierarchical representation of chord model gives around 5~6% higher average accuracy than PCP representation.

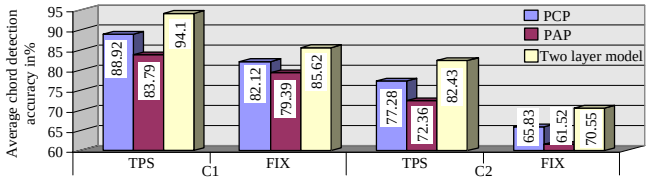


Figure 6: Average detection accuracy of the chords.

Figure 7 shows the average detection accuracy of the four chords types. Experimental results indicate that chord types detection is more challenging than individual chord detection. Laden and Keefe (1974) [5] highlighted that the capability of the PAP approach for chord type detection is more significant than PCP approach. However they haven't presented the statistics to support their claim. Our experiments on C1-TPS reflect that PCP feature gives around 1% of higher average accuracy than with PAP feature. Our proposed 2 layer chord model gives better average chord types detection accuracy for both C1 and C2 data sets.

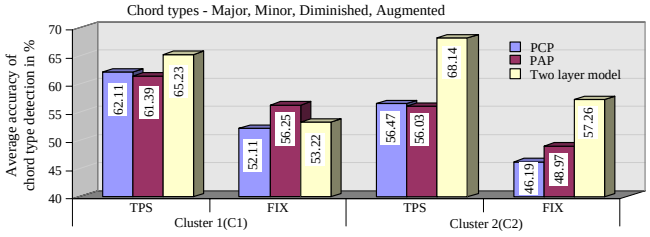


Figure 7: Average accuracy of chord type detection

5. CONCLUSIONS

In this paper we analyze two representations of music chord tonal characteristic and propose a hierarchical chord model for chord detection. Individual octave analysis test reveals that the effects of F0, sub-harmonic, and harmonic of the notes, which comprise the chord, are important for chord detection. For chord detection, pitch class profile approach, which average the tonal effects across the octaves, performs better than psycho-acoustical approach where effects are merged a crossed the octaves. Experimental results revealed that our proposed hierarchical chord model together with tempo proportional signal segmentation can improve both the chord detection and chord type detection.

We find that it will be a continuous effort to explore more effective ways to detect tonal characteristics (fundamental frequencies, sub-harmonics and harmonics) of music notes and to incorporate their effects to improve the chord detection accuracy. We agree that, without a large dataset, neither a solid evaluation of the performances nor a fair benchmarking of algorithm can be established. We plan to expand our cluster 2 dataset (currently 40 songs) in the future and carry out more detailed experiments in this aspect.

6. REFERENCES

[1]. T. Fujishima, "Real time Chord Recognition of Musical Sound: A System Using Lisp Music", In *Proc. ICMC*, Beijing, 1999.

[2]. J. L. Goldstein, "An Optimum Processor Theory for the Central Formation of the Pitch of Complex Tones", In *JASA*, Vol. 54, 1973.

[3]. M. Goto, "A Predominant F0 Estimation Method for CD Recording: MAP Estimation using EM Algorithm for Adaptive Tone Models", In *Proc. Of IEEE ICASSP*, Utah, 2001.

[4]. C. L. Krumhansl, "The Psychological Representation of Music Pitch in a Tonal Context", In *Journal of Cognitive Psychology*, Vol. 11, No. 3. 1979.

[5]. B. Laden, and D. H. Keefe, "The Representation of Pitch in a Neural Net Model of Chord Classification", In *Computer Music Journal*, Vol. 13, No. 4, 1989.

[6]. N. C. Maddage, *Content-Based Music Structure Analysis*, Ph.D Dissertation, Department of Computer Science, National University of Singapore, 2005.

[7]. J. A. Moorer, *On the Segmentation and Analysis of Continuous Musical Sound by Digital Computer*, Ph.D Dissertation, Department of Computer Science, Stanford University, 1975.

[8]. R. J. Ritsma, "Frequency Dominants in the Perception of the Pitch of Complex Sounds", In *JASA*, Vol. 42, No. 1, 1967.

[9]. A. Sheh, and D.P.W. Ellis, "Chord Segmentation and Recognition Using EM-Trained Hidden Markvo Models", In *Proc. Of ISMIR*, Maryland, USA, 2003.

[10]. S.S. Stevens, J. Volkmanm, and E.B. Newman, "A Scale for the Measurement of the Psychological Magnitude of Pitch", In *JASA*, Vol. 8, 1937.

[11]. A. Shenoy, R. Mohapatra and Y. Wang, "Key Detection of Acoustic Music Signals", In *Proc. Of IEEE ICME*, Taiwan, 2004.

[12]. E. Terhardt, "Pitch, Consonance and Harmony", In *JASA*, Vol. 55, No. 5, 1974.

[13]. Y. Zhu, M. S. Kankanhalli, S. Gao. "Music Key Detection for Musical Audio", In *Proc of MMM*, Australia, 2005.

[14]. W. Ward, "Subjective Music Pitch", In *JASA*, Vol. 26, 1954.