Presented at the DLSU Research Congress 2016
De La Salle University, Manila, Philippines
March 7-9, 2016

Simple Audio Processing Approaches for Guitar Chord Distinction

Lance Alcabasa[1*] and Nelson Marcos[2*]
1 Software Technology Department, College of Computer Studies
2 Software Technology Department, College of Computer Studies
**Corresponding Authors: lance.alcabasa@dlsu.edu.ph, nelson.marcos@dlsu.edu.ph*

Abstract: Audio processing has always been one of the most widespread topics of research. This covers many areas, one of which is music analysis and transcription. Music analysis and transcription deals with developing algorithms and processes that allow computers to understand, process, and analyze the digital signals from music playback, and transcribe it into existing musical notations such as a music score. This involves a combination of a variety of modules which include, but not limited to, pitch detection, onset detection, and chord recognition. In particular, working with chords, which are multiple notes played together, presents an interesting topic. Recognizing what chord was played, or chord recognition, is an essential part of music analysis and transcription. Equally as important though should be chord distinction, which is distinguishing whether a particular sound from a music recording should be considered a chord or a single note. However, this problem is not as widely tackled as chord recognition. This paper aims to present two simple approaches aimed at distinguishing chords from single notes, which include working and analyzing the frequency domain of a music recording as well as working with comparison of chroma vectors. The results and analysis are then shown, with the best result achieving around 85% accuracy based on 22 audio files.

Key Words: guitar ; music ; chord ; distinction ; labeling

1. Overview

1.1 Introduction

As music becomes more and more relevant in the world today, so too does research regarding this particular area. Audio processing for music has been one of the interesting areas in research, and includes a wide variety of different research areas. One of these areas is chord detection.

The purpose of chord detection is to be able to detect chords as well as assign chord labels to each chord. (Stark, A. M. & Plumbley, M.D.) As such, chord detection in this aspect can be divided into two sub-categories, chord labelling and chord distinction.

Chord labelling is the aspect of chord detection module that determines which chord label or chord structure is assigned to a particular point in the signal.

On the other hand, chord distinction deals with distinguishing which sounds are single-notes and which sounds are chords.

Both are equally important for chord detection. However, it was observed that there is a lack of previous works tackling chord distinction. As such the focus of this research aims to present simple approaches for chord distinction. (Alcabasa, L. & Marcos, N., 2011)

1.2 Chords Overview

Chords are notes that are played together simultaneously. There are many varieties of chords such as Major, Minor, Major Seventh, Minor Seventh, and Diminished among many others. Each chord has its own structure, which pertains to which notes comprise together the chord, as well as how harmonic or dissonant the chord will sound. (Bernstein, M., 1937)

A chord structure usually begins on the base note or root, which is the note that is the basis for the chord (C in C Major, D# in D# Minor, etc.). Determining the other notes is then as simple as counting from the current note to the next note. For example, a Major chord structure comprises of the root, the note 4 steps from the root, and lastly the note 3 steps from the previous note, which builds into C, E, G to make up the C Major chord. Figure 1 denotes some of the chord structures considered in this research.

Major (Maj) = 0 + 4 + 3

Minor (Min) = 0 + 3 + 4

Maj Seventh (M7) = 0 + 4 + 3 + 4

Min Seventh (m7) = 0 + 3 + 4 + 3

Dominant Seventh (7) = 0 + 4 + 3 + 3

Minor Major Seventh (mM7) = 0 + 3 + 4 + 4

Diminished (Dim) = 0 + 3 + 3

Figure 1. Chord Structures.

2. METHODOLOGY

Because of the lack of previous works regarding this aspect, experimentations and observations were mainly the basis for the approaches discussed below. These include chroma vectors and spectral (or frequency) content of the signal.

2.1 Chroma Vectors

Chroma vectors are typically a size 12 array where in each input represents a tone, starting from C up to B. Chroma bins can be used to represent chords based on their input, which is a numerical value usually between 0 and 1. (Stark, A. M. & Plumbley, M.D.) For example, Figure 2 shows the tone components of a C Major chord.

1	0	0	0	1	0	0	1	0	0	0	0
C	C#	D	D#	E	F	F#	G	G#	A	A#	B

Figure 2: Chroma Bin of C Major Chord

This approach works by calculating and comparing chroma vectors at various points in time. The chroma vector at any given time is computed by the system, allocating the detected frequencies to the appropriate chroma bins. Afterwards, it is compared to the chroma vectors of possible related chords based on its base note (determined by which chroma bin has the highest distribution) using a cosine similarity algorithm.

The theory for working with this approach is that the chroma vector of a chord is more evenly distributed compared to a single-note (which has more intensity focused on a single chroma-bin). In controlled and ideal scenarios (e.g., simulated sin wave, MIDI generated audio), the results supported the theory. However, upon further testing with real-life scenarios (which includes recording through a mic or taking an excerpt from a commercial track), noise was apparent which resulted in inconsistent outcomes, particularly with single-notes being identified as chords instead. Additionally, processing time was a bit long due to multiple calculations and iterations needed in comparing chroma vectors to each other. Thus, an alternative approach was developed.

2.2 Frequency Domain Mean / Variance

An alternative approach was developed and works on using the mean or variance of the spectral content of the signal (or frequency domain) instead of chroma vectors.

The frequency domain, shown in Figure 3, can be achieved by applying the Fast Fourier Transform to the time-domain of the signal.

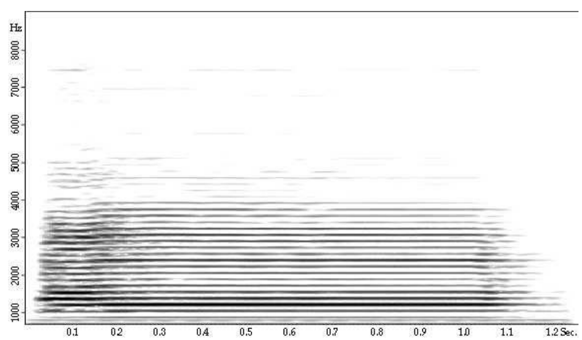

Figure 3. Frequency Domain Representation. (Burk, et al., n.d.)

The frequency domain contains information on the different intensities of various frequencies at different points in time. In the figure, the x-axis represents time, the y-axis represents frequency, and the z-axis (which corresponds to how dark a line is) represents the intensity.

Another representation for the frequency domain is shown in Figure 4. Since FFT works on different blocks of the signal, each block contains information regarding different frequencies and intensities for that block. Each bin corresponds to a frequency range, which is dependent on the block size used by the FFT as well as the sampling rate of the audio.

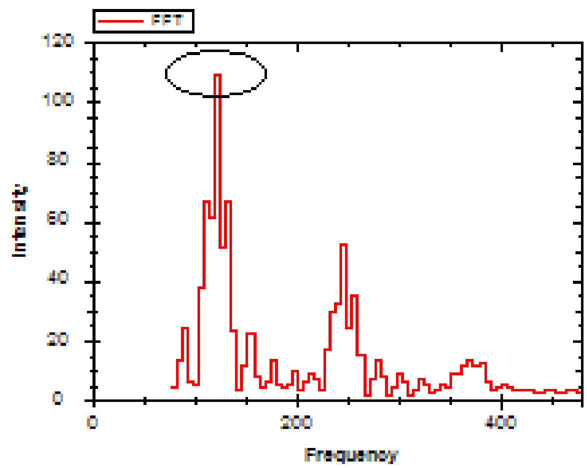

Figure 4. Block Frequency Domain.

The approach works on the theory that the frequency domain of a chord is noisier, thus having higher intensities which results into higher mean and/or variance than the frequency domain of a single-note, as shown in Figure 5 and Figure 6. The algorithms for mean and variance are shown in Formula 1 and Formula 2 respectively, where n is the total frame size and X_i is the current intensity value of the current frequency bin.

$$mean = \frac{1}{n} * \sum_{i=0}^{n} X_i$$

Formula 1. Mean.

$$variance = \frac{1}{n} * \sum_{i=0}^{n} (X_i - mean)^2$$

Formula 2. Variance.

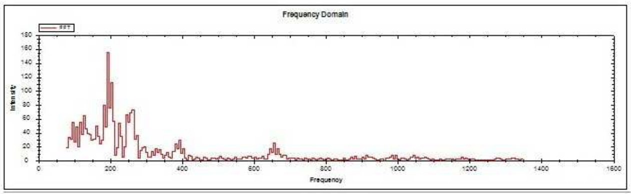

Figure 5. Frequency Domain of a Chord.

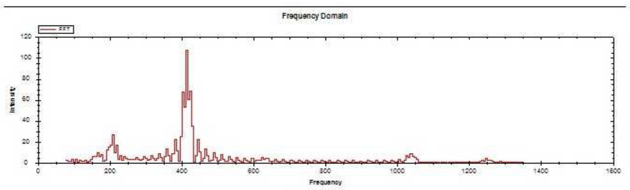

Figure 6. Frequency Domain of a Single-Note.

This approach was observed to work better, able to have more success in distinguishing chords from single-notes, as will be discussed on the next section.

However, both approaches have a limitation in the form of needing an appropriate threshold, which may change depending on the data presented which varies depending on various elements. Additionally, it was also observed that inconsistencies regarding the noisiness of the frequency domain for single-notes were present in some of the audio files (particularly those recorded with a mic).

3. RESULTS AND DISCUSSION

Testing was done on 22 audio files with varying elements. A brief overview of this include 5 songs for basic testing, 3 songs that only contain chords, 6 songs that only contain single-notes, 4 songs that contain both single-notes and chords, and lastly, 4 songs that are excerpts from CD quality files (MP3 files which were translated to .WAV files). 18 audio files were recorded through a microphone with a classical guitar, while the 4 remaining were excerpts from CD recordings (annotated with a "Sample" in the figure). The audio files used went through a pre-processing module which attempted to improve the signal by using a running-pass filter and band-pass filter. Figure 7 shows the data set used which includes how many single-notes and chords are present for each audio sample.

Names	Total Single-Notes	Total Chords
Pitch Test	32	0
Speed Test	30	0
Major Chords	0	7
Minor Chords	0	7
Diminished Chords	0	7
Happy Birthday	26	0
Joy To the World	53	0
London Bridge	24	0
Row Row Your Boat	23	0
Stand By Me	73	0
Twinkle Twinkle	42	0
Let It Be	0	32
Three Little Birds	0	53
21 Guns	0	32
Billy Jean	54	2
Narda	45	54
I Was Only Joking	9	46
Sige	40	16
(Sample) Wake Me Up When September Ends	24	0
(Sample) Stairway To Heaven	26	0
(Sample) Tribute	34	0
(Sample) Unbreakable	46	0

Figure 7. Sample Data Set and Content Detail.

The aim of the test was to determine the accuracy of distinction for three separate categories, which include 1) percentage correct of distinguishing single-notes, 2) percentage correct for distinguishing chords chords, and 3) percentage correct when taking into account both single-notes and chords. Both the Chroma Vectors and Frequency Domain Mean/Variance approach discussed in Section 2.2 were used for testing. Trial and error experimentation was done prior to testing for configuring the thresholds to get the best results possible.

The results for the test using the Chroma Vectors approach are shown in Table 1. It can be observed that the results can be further improved, achieving only around 62% when considering both single-notes and chords. As seen on the table, the approach had difficulty in distinguishing single-notes, incorrectly labeling many as chords instead.

Table 1. Chord Distinction Results using Chroma Vector approach.

Category	Accuracy
% correct labeling for single-notes	68.38%
% correct labeling for chords	86.49%
% correct labeling for both	62.58%

The results for the test using the Frequency Domain Mean/Variance approach are shown in Table 2. Overall, the results are generally better than the previous approach, with over 85% accuracy when considering both single-notes and chords. It should be noted that for this approach, it was able to correctly identify almost all single-notes correctly, and had more difficulty distinguishing chords instead. This is in contrast to the previous approach.

Table 2. Chord Distinction Results using Frequency Domain Mean / Variance approach.

Category	Accuracy
% correct labeling for single-notes	95.03%
% correct labeling for chords	77.58%
% correct labeling for both	86.14%

Lastly, comparisons to commercial software were considered, however, no software was found that dealt specifically with chord distinction.

4. CONCLUSIONS

In general, music has always been an important part of the world and human culture. As such lots of previous research and software have been done to try to allow computer to understand music. This paper tackles one part of music recognition which is chord distinction. Chord distinction is a rare topic, with most other previous works mainly dealing with chord labeling. Chord distinction is important to be able to differentiate single-notes from chords.

This paper has proposed two simple approaches. These work on the theory that single-notes have a frequency domain that has intensities focused mostly on the particular frequency of a single-note, resulting in few harmonics as well. Chords on the other hand, have multiple notes played together, resulting in its frequency domain having a more spread-out nature. This behavior is shown in previous figures.

The problem with the theory is that it is not perfect. There are times in which single-notes also have spread-out frequencies. A possible reason for this is because of string disturbance. Because of the nature of sound, playing a particular note in the guitar may cause some of the other strings to be disturbed and 'ring-out'. While this may not be very noticeable to human hearing, this affects how the signal is formed. As a result, various inconsistencies are then produced, particularly in the frequency domain.

Additionally, working with the approaches required a thresholding function to determine if a specific point in the signal should be qualified as a chord or a single-note. As with all thresholding functions, there are various problems in determining how high or low to set the threshold. Currently, static thresholding was implemented to determine if it is a chord or not, with the value of the threshold set after testing different values on different data.

For future recommendations, improving the existing approaches are suggested, such as making use of dynamic thresholding to try and improve results. Additionally, working with chroma vectors present an interesting approach and integrating other algorithms, approaches, and/or theories to it might be promising. Lastly, further ventures to different approaches are also suggested to improve and widen the field of study regarding chord distinction.

6. REFERENCES

Alcabasa, L., & Marcos, N. (2012). Automatic Guitar Music Transcription. ACSAT International Conference, Nov. 2012, pp. 197 – 202.

Bernstein, M. (1937). An introduction to music. Prentice Hall Inc.

Burk, P., Polansky, L., Repetto, D., Roberts, M., & Rockmore, D. (2011). Music and computers.

Chua, E., Hernandez, R., Lautchang, W., & Ong, R. (2009). Guitar Music Transcription.

Stark, A. M., & Plumbley, M. D. (n.d.). Real-time chord recognition for live performance.