Development of a Novel Method for Automatic Detection of Musical Chords
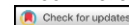
Samin Yaseer Mahmud[1]
Farjana Snigdha[2]
Adnan Siraj Rakin[3]

(Corresponding Author)

[1]*Department of Computer Science and Engineering, Bangladesh University of Engineering & Technology, Bangladesh*
Email: samin.yaseer@gmail.com Tel: +8801672224386
[2]*Department of Electrical and Electronic Engineering, Bangladesh University of Engineering & Technology, Bangladesh*
Email: snigdha152buet@gmail.com Tel: +8801710281183
[3]*Department of Electrical and Electronic Engineering, University of Central Florida, USA*
Email: adnanrakin@knights.ucf.edu Tel: +13219470327

ABSTRACT

This paper represents a method to extract guitar chords from a given audio file using a probabilistic approach called Maximum likelihood estimation. The audio file is split into smaller clips and then it is transformed from time domain into frequency domain using Fourier Transformation. There are multiple known frequencies of musical notes we denote them as reference frequencies. A chord basically is a combination of multiple frequencies. Fourier transformation allows us to identify the frequencies that have precedence over other frequencies in that clip. So, we identify the frequencies having precedence over other frequencies and match them with the reference frequencies to find out which note they belong to. Thus, we get a number of notes in each clip yielding us a specific chord. If we fail to obtain a chord for any sample clip, we follow a probabilistic approach which is termed as 'Maximum Likelihood Estimation' and we use it to approximate the musical chord for the first time with high level of accuracy.

Keywords: *Notes, Chord, Fourier transformation, Maximum likelihood estimation.*

1. INTRODUCTION

Musical chord detection has been a matter of great interest among scientist for quite some time now [1]. The paper is based on some concepts regarding music like Musical Notes, Scale, Octaves, Musical chords. Here brief idea on these terms is included. Musical notes are the atoms of music. Sounds having same pitch are identified as same notes. There are twelve different notes in music. Named as C, C#, D, D#, E, F, F#, G, G#, A, A#, B. Now each of these notes have different frequencies. They are called reference frequencies.

	C	C#	D	Eb	E	F	F#	G	G#	A	Bb	B
0	16.35	17.32	18.35	19.45	20.60	21.83	23.12	24.50	25.96	27.50	29.14	30.87
1	32.70	34.65	36.71	38.89	41.20	43.65	46.25	49.00	51.91	55.00	58.27	61.74
2	65.41	69.30	73.42	77.78	82.41	87.31	92.50	98.00	103.8	110.0	116.5	123.5
3	130.8	138.6	146.8	155.6	164.8	174.6	185.0	196.0	207.7	220.0	233.1	246.9
4	261.6	277.2	293.7	311.1	329.6	349.2	370.0	392.0	415.3	440.0	466.2	493.9
5	523.3	554.4	587.3	622.3	659.3	698.5	740.0	784.0	830.6	880.0	932.3	987.8
6	1047	1109	1175	1245	1319	1397	1480	1568	1661	1760	1865	1976
7	2093	2217	2349	2489	2637	2794	2960	3136	3322	3520	3729	3951
8	4186	4435	4699	4978	5274	5588	5920	6272	6645	7040	7459	7902

Figure-1. Frequencies of Musical Notes

Source: (URL: https://sites.google.com/site/zeroknee/hacks/trackingmacaddress)

The notes appear in a cyclic pattern. That means after C, C#, D up to A, A#, B it will again have C but this time frequency of C will be double of the previous frequency of C. There are many other notes that has frequency which is a power of 2s multiple like 4, 8, 16 of a specific note [2]. Each of these notes belongs to a different octave. So, an octave is consisted of notes in between a frequency and its double or half frequency. A note appearing in a musical chord might belong to multiple octaves simultaneously. The two consecutive notes in the note cycle is termed as Semi Tone apart. That means note C and note C# are a semi tone apart. Two semi tones create a Tone [2]. A scale is a collection of seven individual notes from the note cycle in ascending order of frequency. The structure of a major scale is in the format TTSTTTS. For example, the C Major Scale is consisted of notes C, D, E, F, G, A, B. The musical chord is a combination of multiple notes of a scale playing simultaneously [2]. Usually three notes create a chord and they are called triads. But there could be more like four or five or even six notes of a scale that form a complex chord [2]. The most common chord variations are Major chord and Minor chords. The major chords consist of the 1st, 3rd and 5th note of a major scale [2]. The minor chords consist of the 1st, Flat 3rd, 5th note of the scale.

Automatic chord detection has always been a field of challenge and different methods has been suggested over last few years regarding its solutions [3]. Different approach has been taken to solve this issue and lot of them has brought about good rate of success [4-7]. However, likely hood estimation is an approach that is bound to give better results in this field [8, 9]. Adding probabilistic approach to that would certainly increase the accuracy of the estimation. Here simple probabilistic approaches were used which found to be successful in this field and other fields of work as well [10-14]. By combining these two methods in this paper we have proposed a novel approach in detecting musical chords.

2. DIGITAL SIGNAL PROCESSING METHOLOGY

A. Discrete Fourier Transformation

The audio signal contains data points which represent amplitude of the signal in the time domain. The audio signal is a signal consisted of multiple sinusoidal waves of various frequencies. The signal is:

$X(t) = \sin 2\pi f_1 t + \sin 2\pi f_2 t + \sin 2\pi f_3 t + \ldots\ldots + \sin 2\pi f_n t$
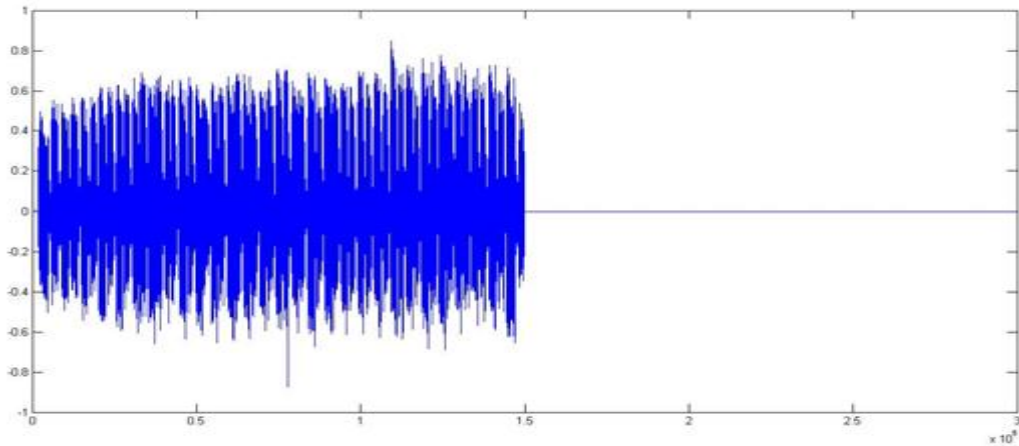

Figure-2.1. Audio signal in Time domain

The individual frequencies of the sinusoidal signals needed to find out as it will help to detect the musical chord. For this Discrete Fourier Transformation [15-17] is performed on sample clip which transforms signal from time domain to frequency domain. Human brain is more responsive to amplitude than phase of a signal.

$X[t] = F[x(t)]$

$X[t] = \sum x(n) e^{-j2\pi tn/N}$

$e^{-j\Theta} = \cos \Theta - j \sin \Theta$

$X[t] = \sum x(n) [\cos(2\pi tn/N) - j \sin(2\pi tn/N)]$

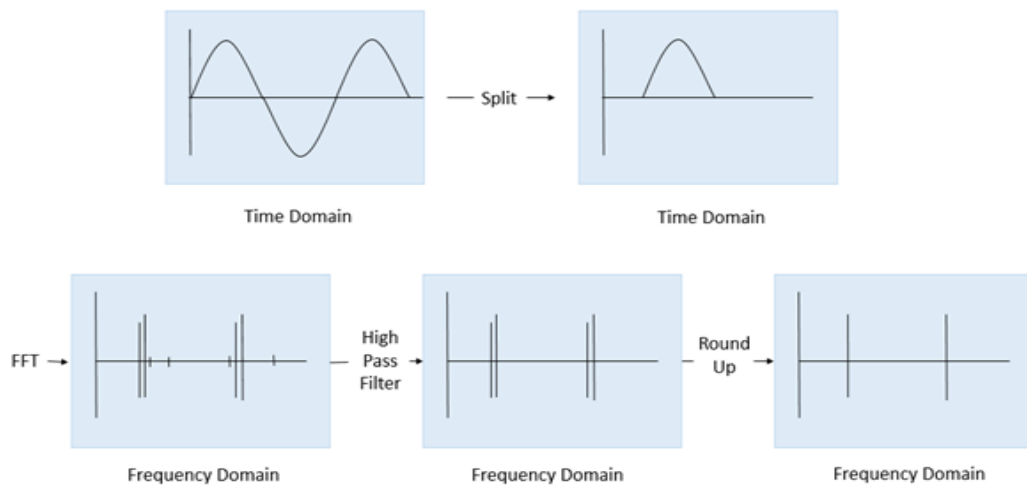**Figure-2.2.** Discrete Fourier Transformation

A java library J Transforms is used to do FFT transformation. As the library only uses real part of any signal thus absolute values of complex value of FFT transformed input signal is used. Thus, frequency spectrum is formed. From observation of frequency spectrum some specific frequencies with very amplitude are found. These are referred as peak frequencies.
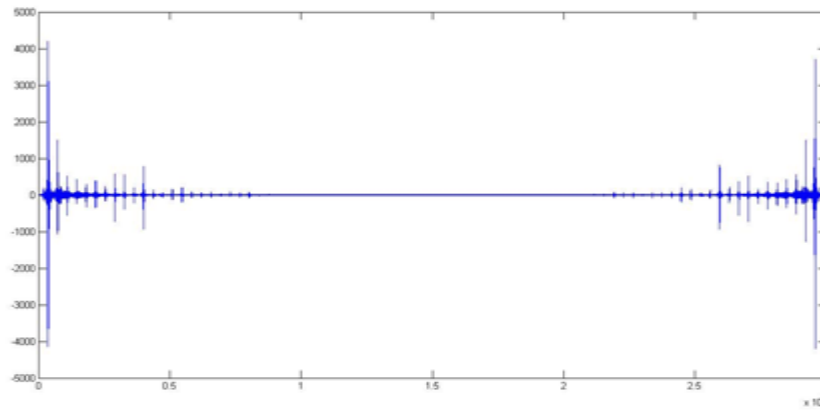
Figure-2.3. Audio signal in frequency domain

B. **High Pass Filtering**

The audio file could have multiple segments that have no chords being played. But some frequencies of noise could be present which can arbitrarily show some chords by calculation. They have very low amplitude values so they are handled by passing the signal through a high pass filter. A threshold value for clipping the silent segments is determined by trial and error on multiple files. The high pass filter is set to that threshold value that only lets the signal having amplitude more than it pass through it and sets the rest to zero.
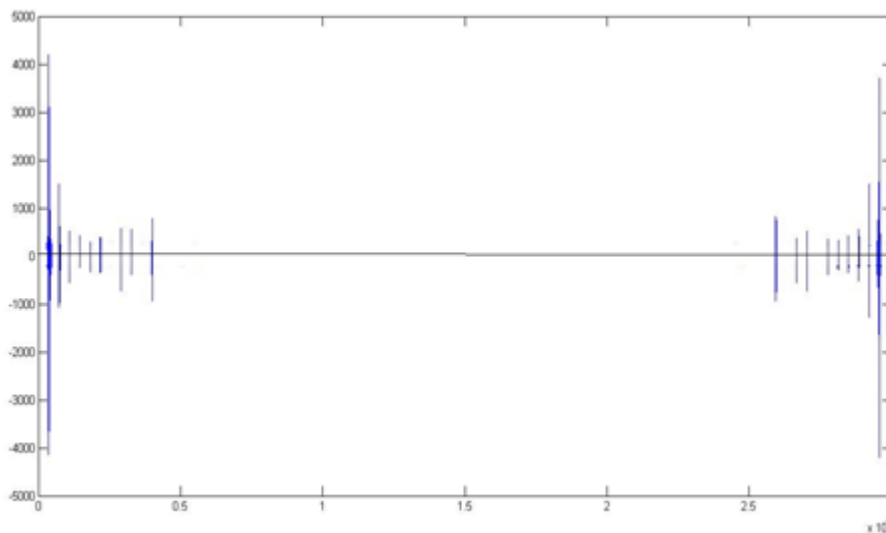
Figure-2.4. Audio signal in frequency domain after High pass filtering

2. DETECTION OF MUSICAL CHORDS

C. **Audio Segmentation and Identifying Notes:**

The audio file is in wav format thus a java library called Simple Wav I/O is used for reading and parsing the data into float values and stored in float array. The array is split into small intervals of around one second so that chord frequency is not altered. The size of the sample clip is:

Size = N * F_s

F_s = Sampling frequency

N = Time interval

D. **Constructing Chord Evidence in Segment:**

The high amplitude peak frequencies are compared with reference frequencies and try to detect the chord. Due to noise frequencies are shifted a bit. That's why around each reference frequency within +-10% frequencies are aggregated and considered as reference frequency. A musical chord is consisted of at least three notes but not all time will contain individual frequency. After getting a specific peak frequency, to which note it belongs to is found. So, an array of notes is formed instead of an array of frequencies to compute the chord. Thus, a set of reference frequencies that represents notes, data set containing combination of notes which construct a chord are formed. The evidence is matched with chord data set to see if it matches to any valid chord. If it matches, that chord is simply assigned to that specific sample clip.

E. **Maximum Likelihood Estimation:**

A probability table is constructed that can qualify the likelihood of observing that particular evidence against each of the 24 candidate chords. This is denoted by P (E | C). The two probabilities are:

P (E_i | C_i) = the probability of observing an event E_i given chord C_i

P (C_i | C_{i-1}) = the probability of seeing chord C_i followed by chord C_{i-1}

Using Bayes rules max P (E, C) is found.

P (E, C) = Π P (E_i | C_i) * Π P (C_i | C_{i-1})

Taking Log on both sides,

log P (E, C) = $\sum$ log P (E_i | C_i) + $\sum$ log P (C_i | C_{i-1})

log P (E_i - C_i) and log P (C_i − C_{i-1}) are two matrices called Evidence probability matrix A and Chord relatedness matrix B respectively.

A [n][i] = log P (E_i | C_i)

B [i-1] [i] = log P (C_i | C_{i-1})

F. **Evidence Probability Matrix:**

The Evidence probability matrix denotes the probability of the evidence occurring given a particular chord is played. The size of this matrix is N * C where N is equal to the number of evidence that is been observed and C is the number of chords in database. From particular evidence it is matched how many notes of that specific chord has match with the evidence and that many points is assigned to that chord. Suppose evidence is C, F, G notes and comparing it with C major chord with is consisted of C, E, G notes. Then there are two notes match between these two. So, two point is assigned to C major. The probability is normalized by dividing a chord's point by total number of points given to all the chords. Similarly, probability for all the chords are computed and Evidence matrix is filled up. Some chords might have zero notes common with a particular evidence. For that zero probability is assigned.

G. **Chord Relatedness Matrix:**

The chord relatedness matrix denotes the probability of one chord appearing given that a particular chord was played right before it. The size of this matrix is N * N where N is the number of chords that are considered. It is found that the appearance of one chord after another is not totally random. The chords maintain some sort of relatedness. These chords are called Related Chords. The related chords of a particular root can be easily found out from the 'Circle of the fifth' table.

In the table, the chords are revolving in circular fashion. If the root of a song is known then its related chords can be found out easily. If a note is a root, its first, fourth and fifth notes of the scale should have its chords as major chords and second, third and sixth note as minor chords. These chords are assigned high points. But it is not impossible for the other chords to occur as there could be exception and change of taste in music. So, other chords are also assigned some points. In this way, the results are normalized and probability matrix of chord relatedness is found.
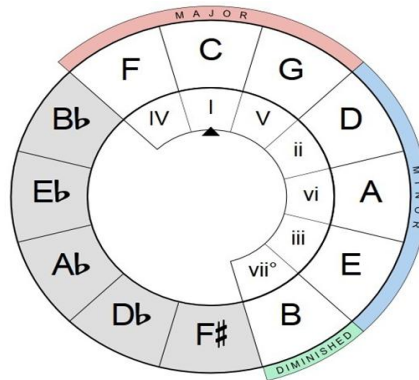

Figure-3.1. The Circle of Fifth

3. EXPERIMENTAL RESULTS

Simulation for detecting guitar chords has been run on several files both consisting of single chords and multiple chords. Both studio recorded and home recorded file has been examined.

For analyzing the performance lets denote three performance matrixes

X = True / positive; that means the chord is played and it is identified

Y = False/ positive; that means a chord is not played but it is identified

Z = False/ negative; that means a chord is played but it is not identified

The ratio of these three factors for various music files is computed.

H. **Studio Recorded Files:**

The chord data base had 12 major and 12 minor chords in total 24 chords. All the major and minor chords were played individually in 3 patterns by acoustic guitar and electric guitar. The studio recorded files were collected from a website called Jam Studio. The total play time of these chords were around 20 minutes. Sampling frequency of the studio files were 44100 Hz. Here is the performance of combination of 10 major and 10 minor chords which were recorded in studio.

The detection performances of studio recorded files were above 90% and chords were easily identified.

Table-1. Performance ratio for Studio File

	Correct Chord		Wrong Chord		Undetected Chord	
Songs	Ratio	%	Ratio	%	Ratio	%
Major Chords	144/150	96%	6/150	4%	0/10	0%
Minor Chords	138/150	92%	12/150	8%	0/10	0%

I. **Home Recorded Files:**

The similar tests were done with some audio files which were recorded by playing an acoustic guitar in a home environment with ambient noise. Due to noise and other issues the accuracy of these audio files was not like the previous one. The accuracy for home recorded files was around 80%. The performance ratios are given below:

Table-2. Performance Ratio for Home Recorded Files

	Correct Chord		**Wrong Chord**		**Undetected Chord**	
Songs	Ratio	%	Ratio	%	Ratio	%
Fall to Pieces	22/30	73%	8/30	27%	0/6	0%
Hotel California	44/54	81%	10/54	19%	1/7	14%
Summer of 69	18/22	81%	6/22	19%	1/5	20%
Nothing else matters	26/31	84%	5/31	16%	0/5	0%
November Rain	36/42	86%	6/42	14%	0/8	0%
All of me	32/37	87%	5/37	13%	0/7	0%
With me	19/26	74%	7/26	26%	1/5	20%

4. CONCLUSION

In the paper the extraction of musical chords from a song is tried. The process is much complicated than it actually look like because when multiple instruments are played it could be impossible to detect chords without separating the instruments sound. So, it is assumed that only one instrument is played at a time. However looking at the accuracy we achieved it is certain that this method would give satisfactory result in a more complex set up than this.

REFERENCES

[1]   M. McVicar, S.-R. Raúl, N. Yizhao, and D. B. Tijl, "Automatic chord estimation from audio: A review of the state of the art," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 22, pp. 556-575, 2014. *View at Google Scholar | View at Publisher*

[2]   T. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," *IEEE Translation Journal on Magnetics in Japan*, vol. 2, pp. 740-741, 1987. *View at Google Scholar | View at Publisher*

[3]   E. Benetos, D. Simon, G. Dimitrios, K. Holger, and K. Anssi, "Automatic music transcription: Challenges and future directions," *Journal of Intelligent Information Systems* vol. 41, pp. 407-434, 2013. *View at Google Scholar | View at Publisher*

[4]   S. Durand, D. Bertrand, and G. Richard, "Enhancing downbeat detection when facing different music styles," presented at the Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference. IEEE, 2014.

[5]   A. J. Eronen, "Evaluation of beats, chords and downbeats from a musical audio signal," *U.S. Patent* vol. 9, pp. 653,056, 2017.

[6]   M. Schedl, M. G. Emilia, and U. Julin, "Music information retrieval: Recent developments and applications," *Foundations and Trends in Information Retrieval*, vol. 8, pp. 127-261, 2014. *View at Google Scholar | View at Publisher*

[7]   A. N. LaCroix, D. F. Alvaro, and R. Corianne, "The relationship between the neural computations for speech and music perception is context-dependent: An activation likelihood estimate study," *Frontiers in Psychology*, vol. 6, p. 1138, 2015. *View at Google Scholar | View at Publisher*

[8]   J. Salamon, G. Emilia, E. P. W. Daniel, and R. Gaël, "Melody extraction from polyphonic music signals: Approaches, applications, and challenges," *IEEE Signal Processing Magazine*, vol. 31, pp. 118-134, 2014. *View at Google Scholar | View at Publisher*

[9]   E. J. Humphrey and P. B. Juan, "Four timely insights on automatic chord estimation," *In ISMIR*, pp. 673-679, 2015.

[10]   A. S. Rakin and S. M. Mominuzzaman, "Finding the chirality of semiconducting DWCNT using empirical equation of radial breathing mode frequency of RRS and optical transition energy," *Journal of Nanoscience and Nanoengineering*, vol. 2, pp. 34-39, 2016. *View at Google Scholar*

[11]   A. S. Rakin, "Developing an empirical equation for the diameter of DWNT and RBM Frequency of RRS," *NanoTrends: A Journal of Nanotechnology and its Applications*, vol. 18, pp. 23–26, 2016. *View at Google Scholar*

[12]   T. Cho and B. P. Juan, "On the relative importance of individual components of chord recognition systems," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 22, pp. 477-492, 2014. *View at Google Scholar | View at Publisher*

[13]   T. Daikoku, Y. Yutaka, and Y. Masato, "Pitch-class distribution modulates the statistical learning of atonal chord sequences," *Brain and Cognition*, vol. 108, pp. 1-10, 2016. *View at Google Scholar | View at Publisher*

[14]   A. S. Rakin and M. M. Sharif, "Double walled carbon nanotube simulator to achieve higher accuracy in finding optical and electrical properties of the tubes," *American Scientific Research Journal for Engineering, Technology, and Sciences*, vol. 26, pp. 282-289, 2016. *View at Google Scholar*

[15]   T. Fujishima, *Realtime chord recognition of musical sound: A system using common lisp music*: In ICMC, 1999.

[16]   S. Fenet, B. Roland, and R. Gaël, "Reassigned time–frequency representations of discrete time signals and application to the constant-Q transform," *Signal Processing*, vol. 132, pp. 170-176, 2017. *View at Google Scholar | View at Publisher*

[17]   K. Wang, A. Ning, N. L. Bing, Z. Yanyong, and L. Lian, "Speech emotion recognition using fourier parameters," *IEEE Transactions on Affective Computing*, vol. 6, pp. 69-75, 2015. *View at Google Scholar*