# MUSIC SIGNAL PROCESSING WITH EMPHASIS ON GENRE CLASSIFICATION

Dissertation submitted towards the fulfillment of requirement for the award of degree of

**MASTER OF ENGINEERING**

in

**ELECTRONICS AND COMMUNICATION ENGINEERING**

Submitted by

**Vaibhav Arora**

**Roll. No. 801261028**

Under the guidance of

**Dr. Ravi Kumar**

**(Assistant Professor)**

Electronics and Communication Engineering Department

THAPAR UNIVERSITY

PATIALA-147004

(July 2014)

# DECLARATION

I hereby declare that the Dissertation entitled **"MUSIC SIGNAL PROCESSING WITH EMPHASIS ON GENRE CLASSIFICATION"** is an authentic record of my own work carried out as requirement for the award of Master of Engineering in Electronics and Communication engineering **at** Thapar University, Patiala under the guidance of **Dr. Ravi Kumar, Assistant Professor, (ECED)**, July 2014.

Date: 17/July/14

Vaibhav Arora
Roll. No. 801261028

It is certified that the above statement is correct to the best of my knowledge and belief.

Date: 17/7/14

Dr. Ravi Kumar
Assistant Professor
Thapar University, Patiala

Countersigned by:

Dr. Sanjay Sharma
Professor and Head (ECED)
Thapar University, Patiala

Dr. S. K. Mohapatra
Dean, Academic Affairs
Thapar University, Patiala

# ACKNOWLEDGEMENTS

First of all, I would like to express my gratitude to **Dr. Ravi Kumar, Assistant Professor,** Electronics and Communication Engineering Department, Thapar University, Patiala for his patient guidance and support throughout this report. I am truly very fortunate to have the opportunity to work with him. I found this guidance to be extremely valuable.

I am also thankful to **Head of the Department, Dr. Sanjay Sharma** as well as our P.G. co-ordinator **Dr. Kulbir Singh, Associate Professor.** Further, I would like to thank entire faculty member, staff of Electronics and Communication Engineering Department, and then friends who devoted their valuable time and helped me in all possible ways towards successful completion of this work. I thank all those who have contributed directly or indirectly to this work.

Lastly, I would like to thank my parents for their years of unyielding love and for constant support and encouragement. They have always wanted the best for me and I admire their determination and sacrifice.

Vaibhav Arora

(801261028)

# ABSTRACT

Distribution estimation of music signals is necessary both for analysis and synthesis tasks. Genre classification is also one of the most fundamental problems in music signal processing. The present work is an effort to understand the probability distribution of music signals with an aim to classify music genres. For this four well known speech distributions viz. Gaussian, Generalized Gamma, Laplacian and Cauchy have been tested as hypotheses. The distribution estimation has been carried out in time domain, DCT domain and wavelet domain. It was observed that skewed Laplacian distribution describes the music samples most accurately with the peakedness of the distribution being correlated with the genre of music. Although Cauchy distribution along with Laplacian has been a good fit for most of the data, it is analytically shown in this work that Laplacian distribution is a better choice for modeling music signals. Genre classification between Metal and Rock genres was achieved with 78% accuracy using wavelet denoising.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# CHAPTER 1
# INTRODUCTION

**M**USIC plays a vital part in human life. Music spans an enormous range of styles from simple, folk songs, to orchestras and other large ensembles, to a minutely constructed piece of electronic music resulting from months of work in a studio. The focus of this study is the application of signal processing techniques to music signals, in particular to the problems of analyzing an existing music signal (such as a musical piece in a collection) to extract a wide variety of information and descriptions that may be important for different kinds of applications. There is a distinct body of techniques and representations that are molded by the particular properties of music audio—such as the pre-eminence of distinct fundamental periodicities (referred to as pitches), the preponderance of overlapping sound sources in musical ensembles (known as polyphony in music literature), the variety of source characteristics (difference in timbre), and the regular hierarchy of temporal structures (beat).

These tools are more or less different from those encountered in other areas of signal processing, even from closely related fields such as speech signal processing. The more closely the signal processing can reflect and exploit the particular properties of the signals at hand, the more successful it will be in extracting relevant information. Musical signals, despite their enormous diversity, do show a number of key properties that result in the techniques of music signal processing. While many techniques are initially borrowed from speech processing or other areas of signal processing, the unique properties and stringent demands of music signals have dictated that simple repurposing is not enough, leading to some inspired and elegant solutions.

## 1.1 Motivation:-

The common musician's problem is that until a certain level of musical achievement it is not possible to play anything meaningful not heard before, and even if it is somehow possible, its efficiency would be improved through reading sheet music. By reading music it becomes evident what notes need to be played, rather than guess what they are.

For example to play a C-sharp or a C-natural, if the ear is not developed enough to know what note it is just by hearing , it becomes difficult to decipher a song while working it. There are, of course, many advanced players who can play through sound alone, but most don't have this skill. The point is, without reading music it is difficult to guess what notes are being played. Even the most advanced programs are unable to makes sheet scores from mp3 files. They can scan sheet music in and can play it on a keyboard as a MIDI, but can't create sheet scores from an existing mp3 file [1]. Another area which needs to be explored is auto-tagging of musical scores. Assigning labels to musical scores have always been done by human beings, efficient auto-tagging techniques need to be explored. Even though much emphasis has been on music production music analysis remains a potential research area beneficial for musicians.

### 1.2 Musical Background:-

A deep understanding of music theory is not required for this report, but the basics of music structure will be used throughout. Surprisingly, knowledge about auditory perception has a very limited role in most music processing systems, but since music exists to be heard, hearing science promises to help our understanding music perception and should therefore help in the analysis of complex signals.

#### 1.2.1 Structure:-

Popular music tunes are generally structured around a sequence of chords [1]. A band might have a pianist playing the chords, a bassist outlining the chords with a bass line, a drummer keeping the beat. When improvising (free playing), a lead player usually plays scales or riffs over the chords and the forms the melody. Players use scales (sets of notes that work with particular chords), and they build their treble lines to fit on the chord changes (with a lot of experimentation on the use of scale).

The chords can also be used to define a set of keys, or roots, for the song. A given key is associated with a scale, and sometimes many successive chords in a song will fit into one key, allowing a lead player to play within one scale for an extended period of time or the piece of music may be organized in a different way, with key changes within a song and the lead player emphasizing these changes.

### 1.2.2 Psychoacoustics:-

The branch of science studying the psychological and physiological responses associated with sound is referred to as Psychoacoustics [2]. It can be categorized as under psychophysics.

Hearing a sensory and perceptual event .When a person hears something, that something arrives at the ear as a mechanical sound wave traveling through the air, but within the ear it is transformed into neural action potentials which then travel to the brain where they are perceived. Hence, in many problems in acoustics, such as for audio processing, it is advantageous to take into account the mechanics of the environment, and the fact that both the ear and the brain are involved in a person's listening experience.

The ear converts sound waveforms into neural stimuli, so certain differences between audio may be imperceptible. Certain data compression techniques, such as MP3, exploit this fact. The ear has a different response to sounds of different intensity levels referred to as loudness. Telephone networks and audio noise reduction systems exploit this fact by nonlinearly compressing data samples before transmission, and then expanding them at reception. Sounds that are close in frequency produce phantom beat notes, or intermodulation distortion products. A brief introduction of the psychoacoustic parameters follows.

Sound is a wave and waves have amplitude which is a measure of energy. The amount of energy a sound has over an area is referred to as intensity. The same sound will sound more intense if heard in a smaller area. Sounds with a higher intensity are generally called louder. Intensity is measured in decibels. The human ear is more sensitive to high sounds, so perceptually they will seem louder than a low noise of the same intensity. Decibels and intensity do not depend on perception. They can be measured with instruments (eg whisper is about 10 decibels and thunder is 100 decibels).

Pitch is what distinguishes between low and high sounds. A singer singing the same note twice, one an octave above the other indicates a difference in pitch.

Pitch is a perceptual quantity which depends on the frequency of a sound wave. Frequency is the number of wavelengths per unit of time. Frequencies are measured in hertz which is equal to one cycle of compression and rarefaction per second. Thunder has a frequency of 50 hertz, while a whistle may have a frequency of 1,000 hertz.

The human ear can perceive frequencies of 20 to 20,000 hertz. Some animals are able to perceive sounds at even higher frequencies for example humans cannot perceive dog whistles, while they can, is because the frequency of the whistle is too high be processed by human ears.

In nature, bats emit ultrasonic waves i.e. use echolocation to help them know where trees are or to find prey. Ultrasonic waves are also used in submarines to navigate under water in regions where it's too dark to perceive targets visually or targets are too far away.

Some sounds seem pleasant while others are unpleasant. A solo violin player sounds very different than a violin player in a symphony, even if the same note is played. A guitar also sounds different than a flute playing the same pitch. This is because they have a different tonal or sound quality. When a source vibrates, it vibrates with many frequency components at the same time. Sound quality depends on the combination of these frequencies.

If a violin is played, the energy from the hand is transferred to the string, causing it to vibrate. When the whole string vibrates, the lowest pitch is heard. This pitch is called the fundamental [1]. The fundamental is only one of many pitches that the string is producing. The parts of the string vibrating at frequencies higher than those of the fundamental are called overtones. Harmonics are produced by the parts vibrating in whole number multiples of the fundamental. A frequency of three times the fundamental will sound two octaves higher and is called the third harmonic. A frequency five times the fundamental will sound three octaves higher and is called the fifth harmonic. The fundamental is also called the first harmonic. The timbre parameter is not mathematically well defined, mostly it is defined as the negative of the other psychoacoustic parameters

i.e. the quantity which is neither pitch, duration nor frequency therefore it cites further study [1].

The short-time Fourier transform (STFT) is the most popular tool for describing the time-varying energy across different frequency bands. STFT visualized as its magnitude is called the spectrogram (as in Fig 1.1).

$$X(t,k) = \sum_{n=0}^{N-1} w(n)x(n + tN/2) \exp\left(-\frac{j2\pi nk}{N}\right) \quad (1.1)$$

Where x is a discrete-time signal obtained by uniform sampling of a waveform at a sampling rate of Fs Hz, using an –N point tapered window w (e.g., Hamming for w(n)=0.54-0.46cos(2πn/N)) and an overlap of half a window length , t is the number of frames, K=N/2 is the index of the last unique frequency value, and thus X(t,k) corresponds to the window beginning at time t.N/2Fs (in seconds) and frequency.

$$fcoef(k) = \left(\frac{k}{N}\right) Fs \quad (1.2)$$

in Hertz (Hz). If Fs=44100 ,N=4096 and window length is of 92.8 ms, a time resolution of 46.4 ms, and frequency resolution of 10.8 Hz is obtained.

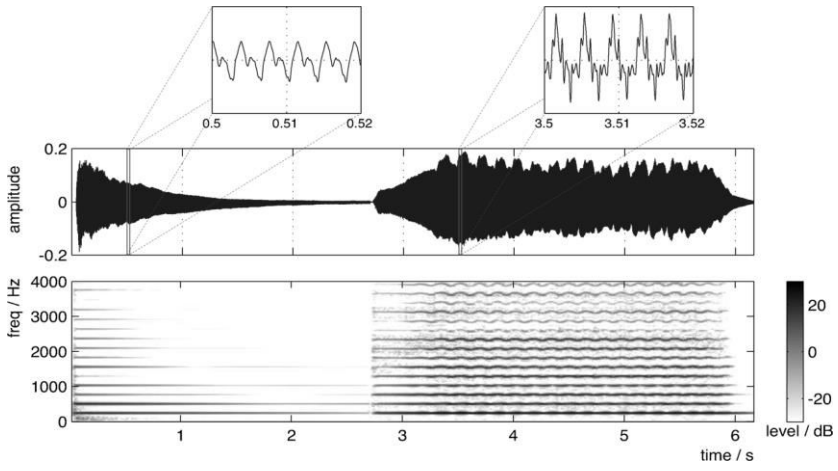


**Figure 1.1 Shows Middle C (262 Hz) played on a piano and a violin. The top pane shows the waveform, with the spectrogram below. Zoomed-in regions shown above the waveform reveal the 3.8-ms fundamental period of both notes [1].**

### 1.2.3 Scales and Chroma [1]:-

Different cultures have developed different musical conventions but, a common feature they all exhibit is the musical "scale," a set of discrete pitches that repeats every octave, from which melodies corresponding to base and treble are constructed. For example, contemporary western music is based on the "equal tempered" scale, which, by a mathematical coincidence, allows the octave to be divided into twelve equal steps on a logarithmic axis preserving intervals corresponding to the most pleasant note combinations.

Dividing the scale as such makes each frequency larger than its predecessor, by an interval known as a semitone. It is possible to divide the octave uniformly into such a small number of steps, and still have these steps give close matches to the simple integer ratios that result in consonant harmonies, e.g. The C major scale spans the octave using seven of the twelve steps—denoted by C, D, E, F, G, A, B. The difference between successive notes is two semitones, except for E/F and B/C which are only one semitone apart.

The "black notes" on a piano are named in reference to the note immediately below, or above, depending on musicological conventions named as a sharp or a flat. The octave degree denoted by these symbols is known as the pitch's chroma and a particular pitch can be specified by assigning a chroma and an octave number (where each numbered octave spans C to B).

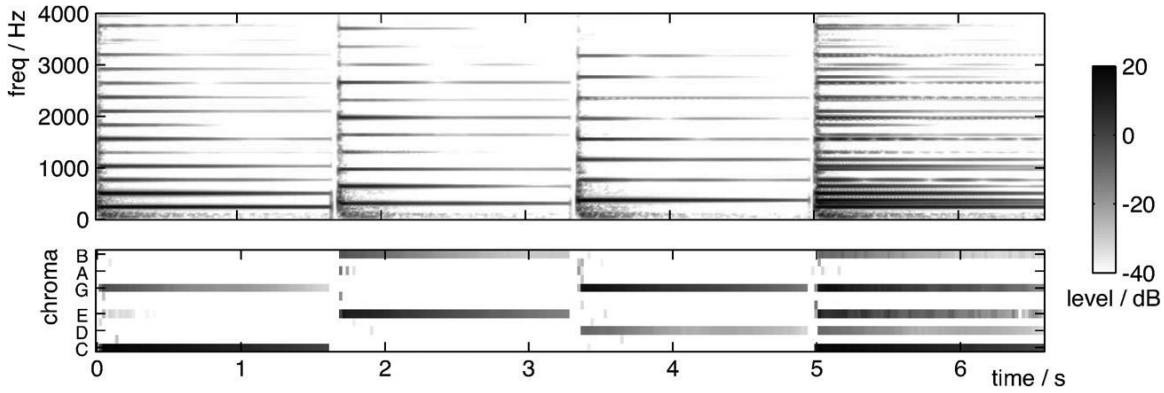


**Fig 1.2 Middle C, followed by the E and G above, then all three notes together—a C Major triad—played on a piano. Top pane shows the spectrogram; bottom pane shows the chroma representation [1].**

### 1.2.4 Harmony

Sequences of pitches create melodies. The “tune” of a musical piece, and the only part reproducible by a monophonic instrument such as the flute, is harmony, the simultaneous presentation of notes at different pitches [1]. Different combinations of notes result in different musical colors or “chords” which remain recognizable regardless of the instrument used to play them.

### 1.2.5 Tempo, beat, and rhythm

The musical aspects of tempo, beat, and rhythm play an important role for the understanding of and the interaction with music. ***Beat*** [3] is the steady pulse that drives music forward and provides the temporal framework of a piece of music. Intuitively, the beat can be described as a sequence of perceived musical pulses that are regularly spaced in time and correspond to the pulse a human taps or nods along when listening to the music.

***Tempo*** [3] refers to the rate of the pulse. Musical pulses typically go along with note onsets or percussive events. Locating such events within a given signal constitutes a fundamental task, which is often referred to as *onset detection*. In this section, an overview of recent approaches for extracting onset, tempo, and beat information from music signals is given and then it is indicated how this information can be applied to derive higher-level rhythmic patterns.

The motive of onset detection is to determine the physical starting times of notes or other musical events as they occur in a music piece. The idea is to capture sudden changes in the music signal, which are caused by the onset of novel/important events. A novelty curve [4] is obtained the peaks of which indicate onset candidates.

There are many different methods for computing novelty curves have. Playing a note on a percussive instrument results in a sudden increase of the signal’s energy which has a pronounced attack phase due to which note onset possibilities may be determined by locating time positions, where the signal’s amplitude envelope starts to increase. The

detection of onsets in the case of non-percussive music, where one often has to deal with soft onsets or blurred note transitions is much more challenging.

It is clearly notable for the case for vocal music or classical music which is dominated by string instruments. Also, in complex polyphonic mixtures, simultaneously occurring events may result in masking effects, which makes it harder to detect individual onsets so more refined methods have to be used for computing the novelty curves like by analyzing the signal's spectral content, pitch, harmony, or phase. A popular approach to onset detection in the frequency domain is the spectral flux where changes of pitch and timbre are detected by analyzing the signal's short-time spectrum.

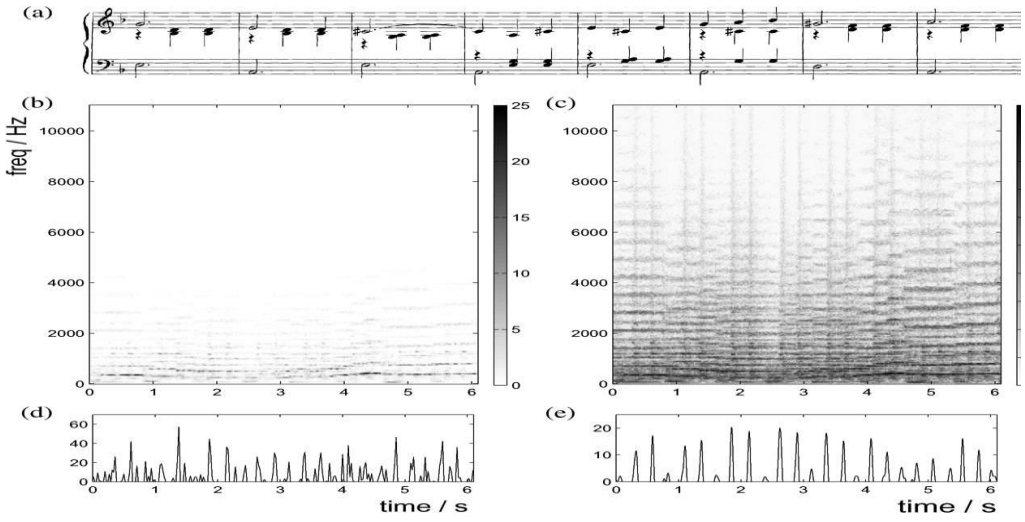


**Figure 1.3 Excerpt of Shostakovich's Waltz No. 2 from the *Suite for Variety Orchestra No.* 1. (a) Score representation (in a piano reduced version). (b) Magnitude spectrogram. (c) Compressed spectrogram using C=1000. (d) Novelty curve derived from (b). (e) Novelty curve derived from (c). [3]A widely used approach to onset detection in the frequency domain is the *spectral flux, where* changes of pitch and timbre are detected by analyzing the signal's short-time spectrum [1].**

## 1.3 Literature Review

**İlker Bayram and Mustafa E. Kamasak, 2013 [6]** propose a simple prior for restoration problems involving oscillatory signals. The prior makes use of underlying analytic frame decomposition with narrow subbands. Other than this, the prior does not

have any other parameters, which makes it simple to use and apply. They demonstrate the utility of the proposed prior through some real audio restoration experiments.

**Bao Guangzhao *et al.,* 2011 [7]** discuss underdetermined blind source separation (BSS) using a compressed sensing (CS) approach, which contains two stages. In the first stage they exploit a modified K-means method to estimate the unknown mixing matrix. The second stage is to separate the sources from the mixed signals using the estimated mixing matrix from the first stage. In the second stage a two-layer sparsity model is used. The two-layer sparsity model assumes that the low frequency components of speech signals are sparse on K-SVD dictionary and the high frequency components are sparse on discrete cosine transformation (DCT) dictionary. This model, taking advantage of two dictionaries, can produce effective separation performance even if the sources are not sparse in time-frequency (TF) domain**.**

**Meinard Müller *et al.*, 2011 [1]** gave an introductory paper describing what all techniques are currently being used to study various psychoacoustic parameters related to music signal processing. They indicated that **Music** signal processing may appear to be the junior relation of the large and mature field of speech signal processing, not least because many techniques and representations originally developed for speech have been applied to music, often with good results. However, music signals possess specific acoustic and structural characteristics that distinguish them from spoken language or other nonmusical signals. This paper provides an overview of some signal analysis techniques that specifically address musical dimensions such as melody, harmony, rhythm, and timbre. They demonstrated that, to be successful, music audio signal processing techniques must be informed by a deep and thorough insight into the nature of music itself.

**Joe Cheri Ross *et al*., 2013 [8]** consider the segmentation of selected melodic motifs from audio signals by computing similarity measures on time series of automatically detected pitch values. The methods are investigated in the context of detecting the signature phrase of Hindustani vocal music compositions (*bandish*) within and across performances.

**Gopala K. Koduri *et al.*, 2013 [9]** approach the description of intonation from a computational perspective, obtaining a compact representation of the pitch track of a recording. First, they extract pitch contours from automatically selected voice segments. Then, they obtain a pitch histogram of its full pitch-range, normalized by the tonic frequency, from which each prominent peak is automatically labeled and parameterized. They validate such parameterization by considering an explorative classification task: three ragas are disambiguated using the characterization of a single track.

**Y.Ueda *et al.*, 2010 [10]** discuss an HMM-based method for detecting chord sequence from musical acoustic signals using percussion-suppressed, Fourier-transformed chroma and delta-chroma features. To reduce the interference often caused by percussive sounds in popular music, they use Harmonic/Percussive Sound Separation (HPSS) technique to suppress percussive sounds and to emphasize harmonic sound components. They also use the Fourier transform of chroma to approximately diagonalize the covariance matrix of feature parameters so as to reduce the number of model parameters without degrading performance. It is shown that HMM with the new features yields higher recognition rates than that with conventional features.

**Matthias Mauch and Simon Dixon, 2010 [11]** present a new method for chord and local key estimation where the analysis of chord sequence and key changes are performed simultaneously. A multi-scale approach for chroma vectors is proposed, and they show an increase in accuracy when the chords are selected from different sized chromas. While the key estimation performs better than a direct template-based method, the chord accuracy shows improved results.

**H. Papadopoulos and G. Peeters, 2008 [12]** in *Proc*. IEEE Int. Conf. 2008 showed that the extraction of tempo and beat information from music recordings constitutes a challenging task in particular for non-percussive music with soft note onsets and time-varying tempo. In this paper, they introduce a novel mid-level representation that captures musically meaningful local pulse information even for the case of complex music. Their main idea is to derive for each time position a sinusoidal kernel that best

explains the local periodic nature of a previously extracted note onset representation. Then they employ an overlap-add technique accumulating all these kernels over time to obtain a single function that reveals the predominant local pulse (PLP).

**Andre Holzapfel *et al.*, 2008 [13]** introduce a novel approach to estimate onsets in musical signals based on the phase spectrum and specifically using the average of the group delay function. A frame-by-frame analysis of a music signal provides the evolution of group delay over time, referred to as phase slope function. Onsets are then detected simply by locating the positive zero-crossings of the phase slope function. The proposed approach is compared to an amplitude-based onset detection approach in the framework of a state-of-the-art system for beat tracking. On a data set of music with less percussive content, the beat tracking accuracy achieved by the system is improved by 82% when the suggested phase-based onset detection approach is used instead of the amplitude based approach, while on a set of music with stronger percussive characteristics both onset detection approaches provide comparable results of accuracy.

**Miguel Alonso *et al.*, 2004 [14]** In this paper the authors have presented an efficient beat tracking algorithm that processes audio recordings. They have also defined the concept of spectral energy flux and used it to derive a new and effective onset detector based on the STFT, an efficient differentiator filter and dynamic thresholding using a median filter. This onset detector displays high performance for a large range of audio signals. In addition, the proposed a tempo tracking system is straightforward to implement and has a relatively low computational cost.

**Goto Masataka, 2010 [15]** described the beat-tracking problem in dealing with real-world audio signals, a beat-tracking model that is a solution to that problem, and applications based on a real-time beat-tracking system. Their experimental results show that the system can recognize the hierarchical beat structure comprising the quarter-note, half-note, and measure levels in audio signals of compact disc recordings.

**S.Dubnov and Naftali tishby, 1995 [16]** provide a simple model that an accounts for timbre properties due to micro fluctuations in the harmonics of the signal. When considering a sound model that consists of an excitation signal passing through a

resonator filter, they find by means of higher order statistical analysis of the excitation, a grouping of sounds according to a common instrumental families of string, woodwind and brass sounds. For resynthesis purposes they model the excitation by family of stochastic, pulse train like functions whose statistical properties resemble those found in real signals. By introducing an idea of "effective number of harmonics" that represents the number of coupled or statistically dependant harmonics among the complete set of partials present in the signal, they show that this number can be calculated directly from the 3$^{rd}$ and fourth moment of the residual. Musically speaking they suggest that micro fluctuations administer a sense of texture within timbre and these texture properties depend upon the concurrence/non concurrence parameter deviations caused by the jitter.

### 1.4 Novel aspects of this Study:-

This study involved in depth analysis of the probability distribution of music signals and its use in genre classification as summarized below:-

1) The probability distribution estimation of music signals in time, DCT and wavelet domains.
2) Genre classification between Metal and Rock Genre using ANNs
3) Comparing the classification accuracy with those obtained by wavelet denoising.

### 1.5 Organization of the Dissertation

As explained in the previous sections, the aim of the work is to study the probability distribution of music signals in time, frequency and wavelet domains. It also aims to study genre classification between Metal and Rock Genres using ANN. In order to achieve the same, the dissertation has been divided into the following chapters:

Chapter 2 describes the probability distribution estimation of music signals in time frequency and Wavelet domains.

Chapter 3 describes the GUI developed during this dissertation.

Chapter 4 outlines genre classification of Metal and Rock genres with wavelet denoising as a preprocessing step.

Finally, Chapter 5 sums up the conclusions of the work and suggests some ideas for the future work.

# Chapter 2

# Distribution Estimation of Music Signals in Time, Frequency, and Wavelet Domains

## 2.1 INTRODUCTION

In signal processing domain music signals are conventionally treated no differently from typical audio signals. However, music signals posses certain acoustic and structural characteristics that distinguish them from speech signals [1]. Music signals can also be classified into a number of genres based upon their psychoacoustic properties. There are several distinctive properties of music signals that necessitate the development of novel signal processing techniques especially suited to them. Distribution estimation of music signals is necessary both for analysis and synthesis tasks [17].

In this chapter the author has attempted to estimate the Probability Density Function (pdf) of music signals belonging to different genres. Music signals are different from speech signals in many ways. First of all, speech is generated by a rather uniform source, i.e. the human vocal tract which is similar in length for all human beings [2], in this way the process of speech generation is governed by several fundamental laws of acoustics which are independent of the language spoken. Furthermore, the adult vocal tract generates a rather limited set of outputs making the spectra of speech signals well defined and predictable. On the other hand, music signals are generated from a varying number of sources whose outputs are highly variable spanning a large band of frequencies. In most of the cases the output of the music signal is not 'speech–like' [3].like speech, music signals can be polyphonic with two or more simultaneous lines of independent melody accompanied by percussion instruments. There is also a remarkable difference in intensities for speech and music signals. Normal intensity speech range from 50db to 75db and shouted speech may reach 83db with peaks and valleys in the range of ±12db. However, music signals can be of the order of 100db with peaks and valleys in the range of ±18db. All this combined together make it necessary to model music signals for processing, synthesis, coding and recognition tasks. This in turn, makes it important to

have a reasonable approximation for the Probability Density Function of music signals. A lot of work has been reported earlier on the estimation of speech signal pdf. The earliest work on this topic can be attributed to Davenport [17] in 1950 in which both male and female speech sample had been investigated in the time domain. Later on novel models for estimation were proposed [18]. The results obtained varied from γ-distribution to Laplacian distribution. However, from application point of view and to ensure computational convenience many papers have assumed speech signal to be Gaussian in nature. However, the nature of pdf changes according to the number of samples taken. It has been shown that samples of a band limited speech signal can be represented by a multivariate Gaussian distribution with a slowly time varying power if samples are taken with short time intervals of less than 5ms [7,19]. A landmark paper by Gazor and Zhang investigated the speech signals in decorrelated domains to approximate the multivariate distribution [19]. It was observed that the distribution shape in both time and frequency domains remained Laplacian. In all the works cited above, a single speech signal has been investigated. Moreover, speech signals typically have intermittent silence intervals which have an effect on properties of the distribution. Musical signals however rarely have a silent time window. Thus, their distribution can be expected to deviate from that of the ordinary speech. To the best of the author's knowledge no work has yet been undertaken to approximate the pdf of the music signals. Due to the above mentioned properties of music signals and the existence of a number of genres with their own set of peculiarities. It has served as a motivation for us to investigate the pdf of music signals by taking samples from several genres of music.

This chapter is organized into the following five sections. The first section describes the music signals used in this studyand their grouping into a particular genre. In the same sectionresults from the time domain estimation has been shown. Two subsequent sections summarize the estimation results in DCT and wavelet domains respectively. To get the best samples in the wavelet domain Multiscale Principal Component Analysis (MSPCA) is applied which has been briefly described in this section. The next section gives an analytical justification for our decision in favor of skewed Laplacian distribution. Finally the conclusions are summarized.

**2.2 Estimation In The Time Domain:-** The present section describes estimation in the time domain.

### 2.2.1. Data Description

The present study has been done on several pieces of music belonging to different genres. The term 'genre' has been used here in a rather loose sense based upon the popular perception.Three recordings/songs each of the following genres have been studied for the PDF estimation task:

- Afro-American folk (Jazz)
- Afro-American folk (Blues)
- East Indian Devotional (Classical)
- East Indian Folk (Punjabi)
- East Indian Pop
- Western Instrumental (Guitar)
- Percussion.

Each musical signal is primarily characterized by its rhythm, which in turn is defined by beats accompanying it. Beats per minute or *tempo* has thus been estimated using MIR toolbox [20] and used as a marker for each song. *Tempo* estimation helps to classify the song into some most easily recognizable classes which are defined conventionally. The following labels have been attached traditionally to a particular tempo based upon beats per minute (BPM).

- Moderato – moderate (86–97 BPM).
- Accelerando - gradually accelerating
- Allegretto – moderately fast (98–109 BPM)
- Allegro – fast, quickly and bright (109–132 BPM)
- Vivace – lively and fast (132–140 BPM)
- Vivacissimo – very fast and lively (140–150 BPM)
- Allegrissimo- very fast (150–167 BPM)

Figure 2.1 shows the time domain amplitude plot of a percussion recording. Beat position location in a signal waveform gives a plethora of information about the tempo which is depicted in the form of onset curve in Figure 2.2 with the beat position shown in the form of red circles.

Table 2.1 enlists the recordings/songs used in this study classified according to genre and tempo. Each recording was divided into blocks of 10ms and the samples (sampled from 10ms blocks with frequency 44100 hz) in the time domain were subjected to Chi-squared, Anderson Darling, and Kolmogorov-Smirnov tests. For the above mentioned tests, four distributions viz. Normal, Laplacian, Cauchy, and Generalized Gamma were chosen as null hypotheses.

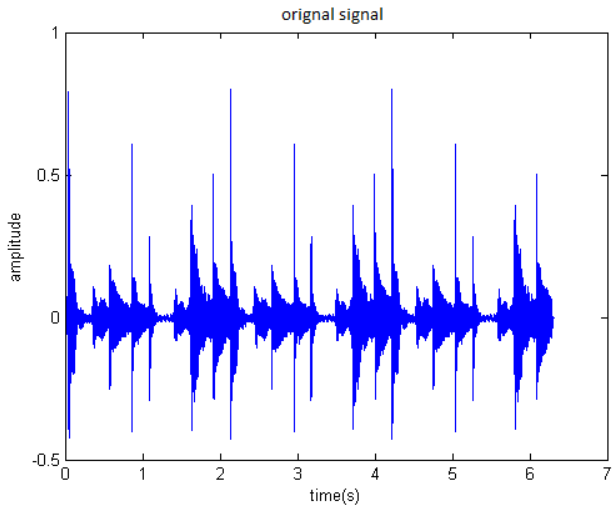


**Figure 2.1 The time domain amplitude plot of a percussion recording.**

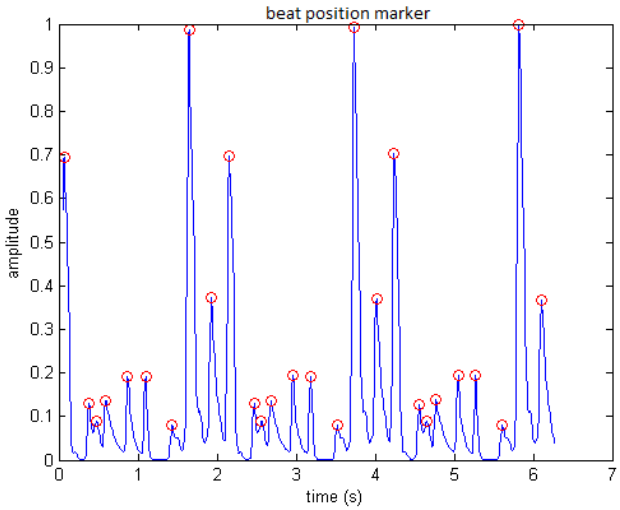


**Figure 2.2 Beat position location in a signal waveform.**

**Table 2.1 Tempo and genre of songs used for this study**

| Genre | Tempo (bpm) | | Time domain | | DCT domain | |
|---|---|---|---|---|---|---|
| | | | Kurtosis | Skew | Kurtosis | Skew |
| Jazz | | | | | | |
| Samp. 1 | 122 | Allegro | 3.32 | -0.30 | 75.32 | -1.25 |
| Samp. 2 | 119 | Allegro | 3.04 | -0.34 | 118.08 | -2.89 |
| Samp. 3 | 142 | Vivacissimo | 3.33 | 0.01 | 34.13 | 0.21 |
| Western Instrumental (Guitar) | | | | | | |
| Samp. 1 | 160 | Allegrissimo | 36.46 | -0.51 | 24.76 | 0.12 |
| Samp. 2 | 164 | Allegrissimo | 27.99 | -0.55 | 7.81 | 0.10 |
| Samp. 3 | 121 | Allegro | 43.88 | -0.60 | 9.10 | -0.01 |
| East Indian Folk (Punjabi) | | | | | | |
| Samp. 1 | 99 | Allegretto | 5.06 | -0.31 | 6.31 | -0.08 |
| Samp. 2 | 187 | Prestissimo | 5.56 | 0.02 | 9.86 | -0.06 |
| Samp. 3 | 102 | Allegretto | 5.41 | -0.09 | 69.10 | -0.71 |

| East Indian Devotional (Classical) | | | | | |
|---|---|---|---|---|---|
| Samp. 1 | 130 Allegro | 5.82 | 0.99 | 90.52 | 1.12 |
| Samp. 2 | 153 Allegrissimo | 4.91 | -0.08 | 306.35 | -0.72 |
| Samp. 3 | 160 Allegrissimo | 5.11 | -1.60 | 350.32 | 2.11 |
| Blues | | | | | |
| Samp. 1 | 90 Moderato | 6.81 | 0.87 | 17.97 | -0.22 |
| Samp. 2 | 63 Adagietto | 5.73 | 0.34 | 17.92 | 0.44 |
| Samp. 3 | 67 Adagietto | 5.39 | 0.36 | 45.47 | -1.57 |
| East Indian Pop. | | | | | |
| Samp. 1 | 190 Prestissimo | 4.58 | -0.07 | 45.81 | -1.35 |
| Samp. 2 | 130 Allegro | 5.47 | 0.21 | 117.60 | 0.23 |
| Samp. 3 | 105 Allegro | 3.98 | 0.17 | 344.48 | 1.32 |
| Percussion | | | | | |
| Samp. 1 | 118 Allegro | 5.06 | 0.05 | 157.19 | -0.79 |
| Samp. 2 | 73 Andante | 5.28 | 0.14 | 591.22 | -3.00 |
| Samp. 3 | 138 Vivace | 2.64 | 0.03 | 1467.13 | -12.88 |

It was observed that for all the genres with 3 recordings each, Cauchy and Laplacian have been the best fitting distributions. The histogram plots have been depicted in Figs. 2.3(a)-2.3(g).

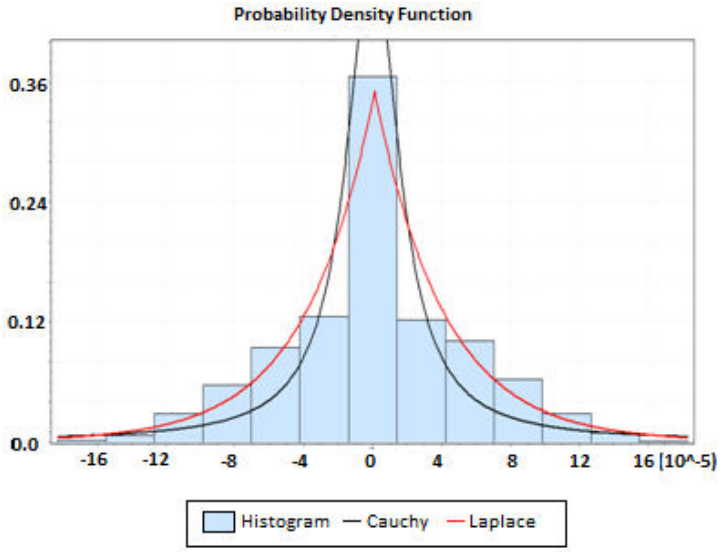


**Figure 2.3(a) Time domain histogram of Jazz.**

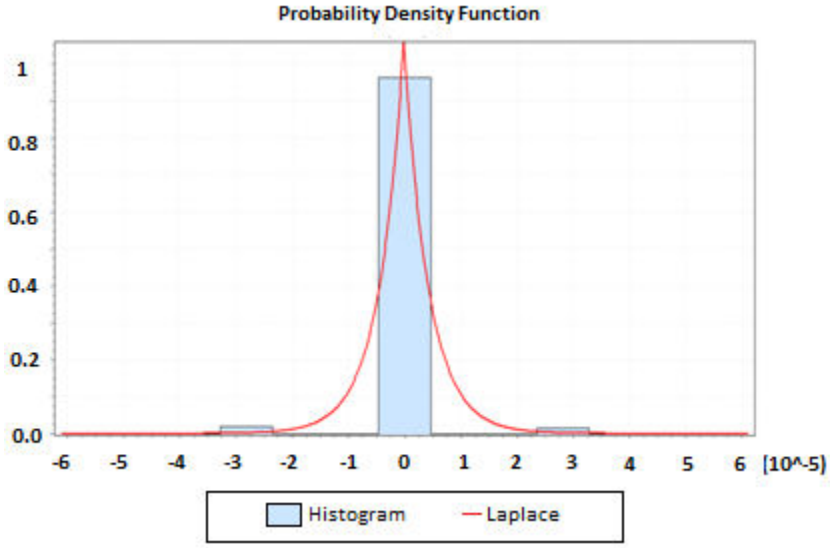


**Figure 2.3(b) Time domain histogram of Western Instrumental (Guitar).**

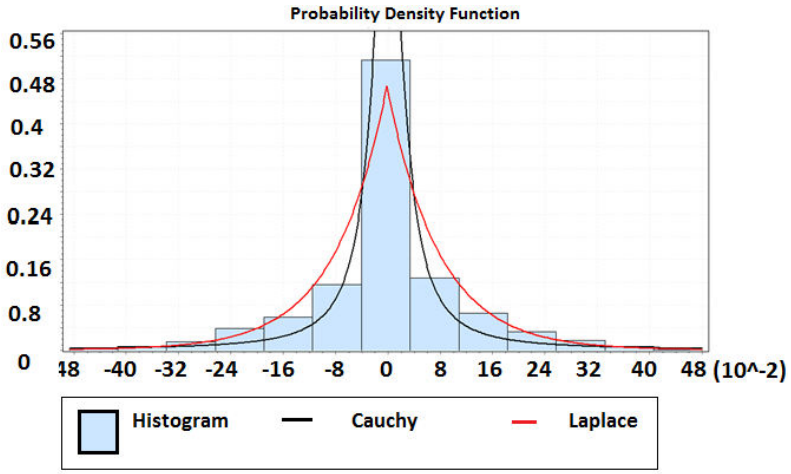


**Figure 2.3(c) Time domain histogram of East Indian Folk (Punjabi)**

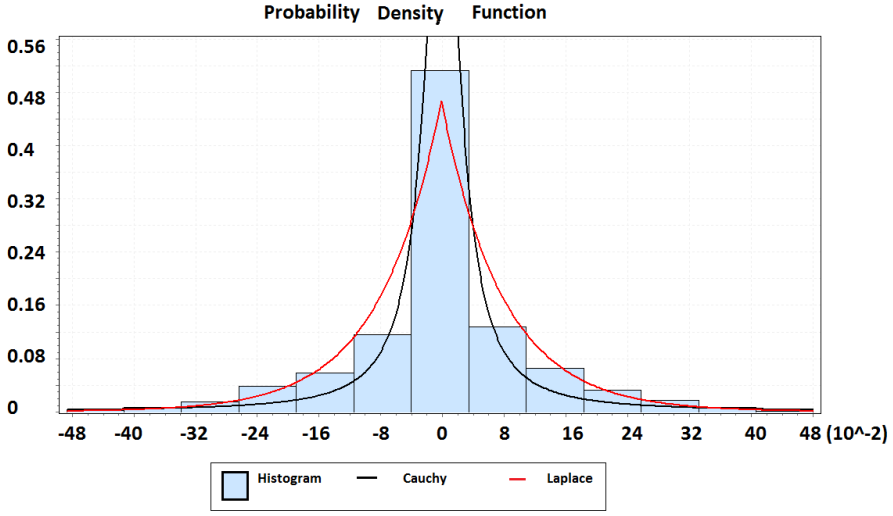


**Figure 2.3(d) Time domain histogram of East Indian Devotional (Classical)**

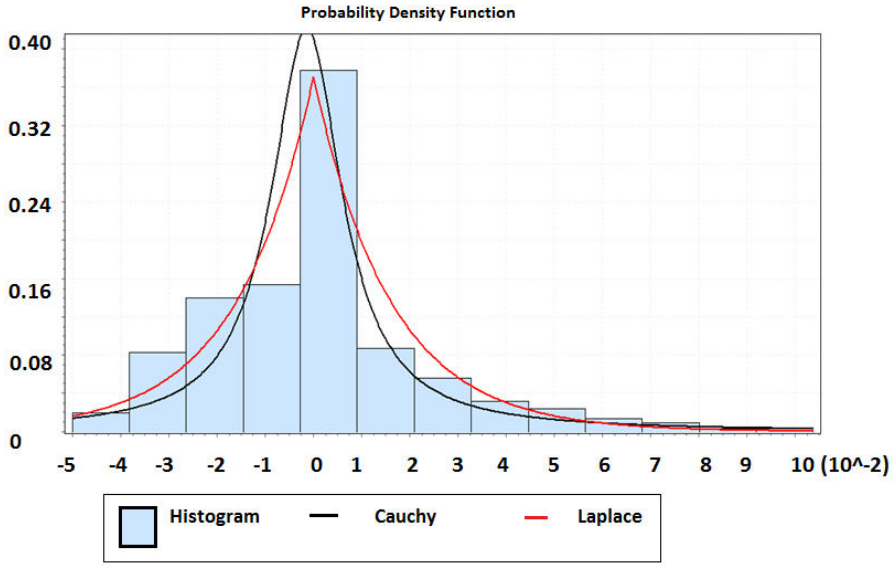


**Figure 2.3(e) Time domain histogram of Blues.**

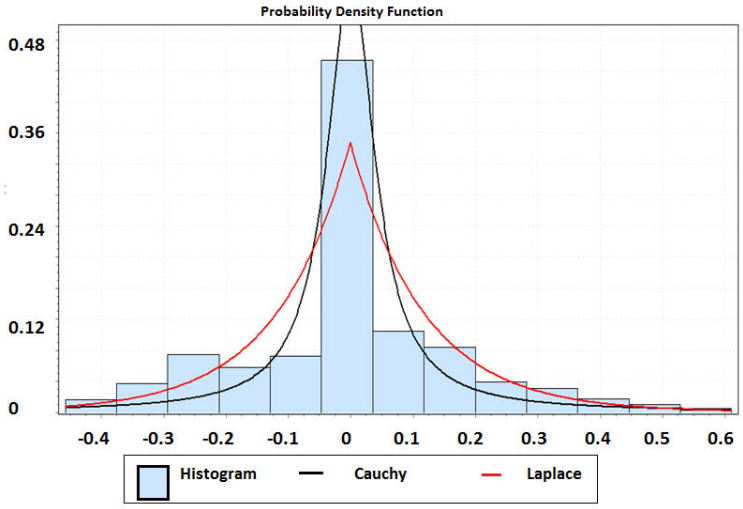


**Figure 2.3(f) Time domain histogram of East Indian Pop.**

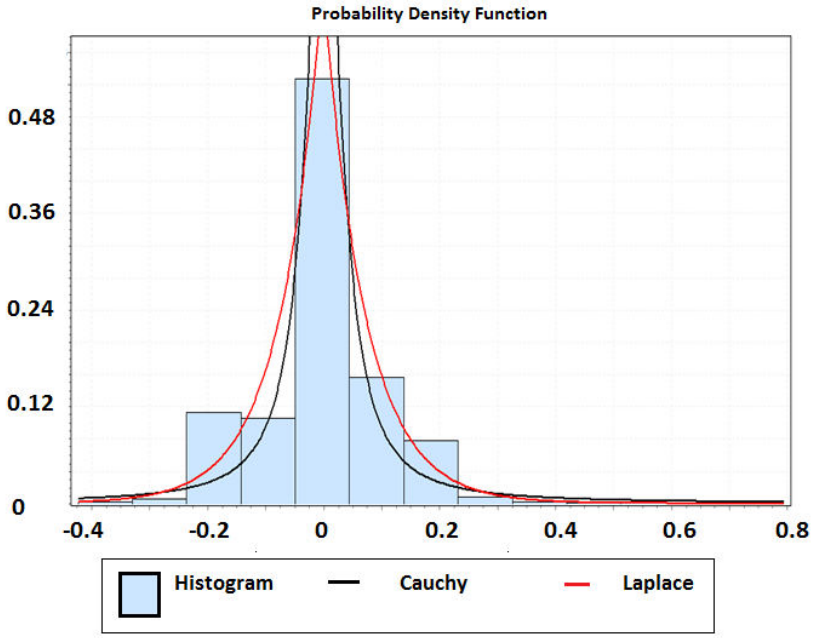


**Figure 2.3(g) Time domain histogram of percussion.**

Both the distributions fitted the data slightly better than one another. The results are shown in Table 2.2 in which the coefficient of fit is shown against a particular genre and a particular test. The results agree with the earlier tests performed on speech signals as far as goodness of fit of Laplacian Distribution is concerned [19]. However, a better fitting Cauchy distribution in some cases can be a good alternative for modeling musical signals.

As a second step to the analysis kurtosis and skew for all the recordings is calculated to ascertain the peakedness and asymmetry of the underlying distributions respectively. Table 2.3 shows the kurtosis and skew of each recording grouped according to genre and tempo. The results reveal an important correlation between the genre of music and the kurtosis where the Guitar recordings have shown an exceptionally high kurtosis. However, recordings belonging to the genres of the super class "East Indian music" exhibit a consistent kurtosis between 4 and 5. Furthermore, all the *jazz* recordings have a normal-like kurtosis (around 3) and all the *blues* recordings exhibit consistent kurtosis between 5 and 6 making them Laplacian like. Most of the recordings exhibit a negative skew.

Although Cauchy distribution has fitted the data well in many cases, following are some issues associated with it that limit its wider acceptability as a preferred choice for modeling tasks. (i) The sample mean is not a good estimator of the location parameter. (ii) It is rather difficult to find the maximum likelihood estimate. (iii) The Cauchy distribution does not have well defined moments. The point (iii) can be explained as follows:

The pdf of Cauchy distribution is given by:

$$P(x) = \frac{1}{\pi}.\frac{1}{1+x^2} \tag{2.1}$$

Whose expectation is

$$E(x) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{1}{1+x^2}\,dx \tag{2.2}$$

The integral in (2) is not completely convergent. Hence, expectation value of Cauchy distribution is undefined. Furthermore, the characteristic function for the Cauchy distribution is given by

$$\varphi(t) = \int_{-\infty}^{\infty} e^{-jtx}\,P(x)dx = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{costx+jsintx}{1+x^2} \tag{2.3}$$

Therefore

$$\varphi(t) = \frac{1}{\pi}\left(\int_0^{\infty}\frac{costx}{1+x^2}dx + \int_{-\infty}^{0}\frac{costx}{1+x^2}dx + \int_0^{\infty}\frac{jsintx}{1+x^2}dx\right) = \frac{2}{\pi}\int_0^{\infty}\frac{costx}{1+x^2}dx = e^{-|t|} \tag{2.4}$$

Since, the characteristic function has no derivative at *t=0*, the distribution has no moments.

Thus, it can be concluded that skewed Laplacian distribution could be a better estimate of musical signal samples in time domain.

**Table 2.2 Kurtosis and Goodness of fit Tests statistic (time domain)**

| Sr no. | Distribution | Kolmogoroov Smirnov (Test statistic) | Anderson Darling (Test statistic) | Chi-Squared(Test statistic) |
|---|---|---|---|---|
| Jazz | | | | |
| 1 | Cauchy | 0.18866 | 169.75 | 1560.7 |
| 2 | Laplace | 0.19155 | 105.6 | 533.7 |
| Western Instrumental (Guitar) | | | | |
| 1 | Cauchy | No fit | No fit | No fit |
| 2 | Laplace | 0.49634 | 1507.9 | 4491.0 |
| East Indian Folk (Punjabi) | | | | |
| 1 | Cauchy | 0.15285 | 103.78 | 5305.1 |
| 2 | Laplace | 0.16381 | 106.31 | 4962.0 |
| East Indian Pop | | | | |
| 1 | Cauchy | 0.16058 | 109.5 | 5043.9 |
| 2 | Laplace | 0.16563 | 96.699 | 4677.5 |
| Blues | | | | |
| 1 | Cauchy | 0.09494 | 53.036 | 1562.5 |
| 2 | Laplace | 0.14847 | 64.385 | 2556.0 |
| East Indian Devotional (Classical) | | | | |
| 1 | Cauchy | 0.18866 | 169.75 | 1560.7 |
| 2 | Laplace | 0.19155 | 105.6 | 533.7 |
| Percussion | | | | |
| 1 | Cauchy | 0.12596 | 73.403 | 3325.0 |
| 2 | Laplace | 0.13702 | 70.146 | 3290.6 |

## 2.3 Estimation In The DCT Domain

As evident from estimation carried out in time domain depicted in Figs. 2.3(a)-2.3(g), the nature of distribution is heavy tailed Laplacian with a high kurtosis and a little skew in most of the cases.

Since, most of the information content of the signal is supposed to be concentrated in tails, corresponding to lower frequencies, it is required that only significant transform components be utilized for the study. Discrete Cosine Transform (DCT) based techniques have previously been applied for speech quantization, compression and noise mitigation [21, 22].

In the context of music, digital watermarking is necessary to accomplish copyright protection and authentication. Since, most of the watermarks are embedded into middle and high frequency coefficients of the DCT, therefore, it is important to have an idea about the underlying distribution of DCT coefficients of musical signals. The same three tests as reported in the previous section have now been performed on DCT coefficients of each recording.

Histogram plots of Fig 2.4(a)-2.4(g) have been fitted with Cauchy and Laplace Distributions since, these distributions have again given best goodness of fit results as evident from fitness coefficient values shown in Table 2.3.

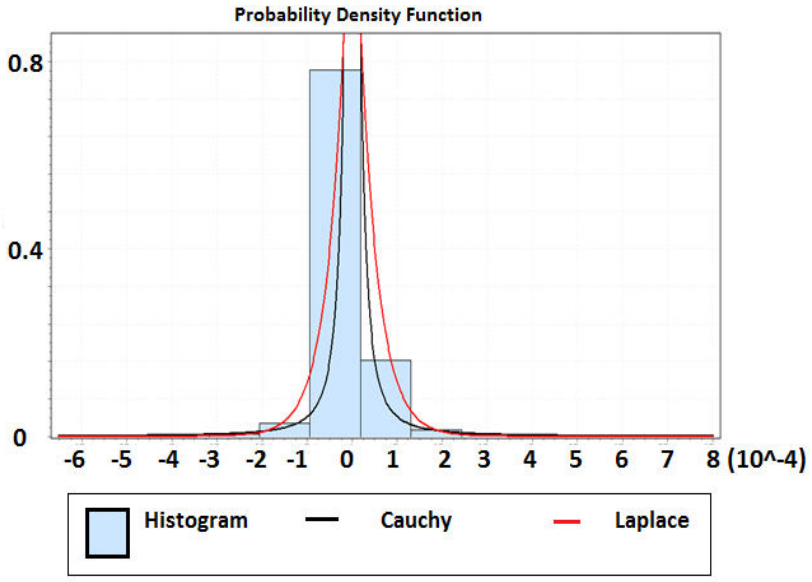


**Figure 2.4(a) DCT domain histogram of Jazz.**

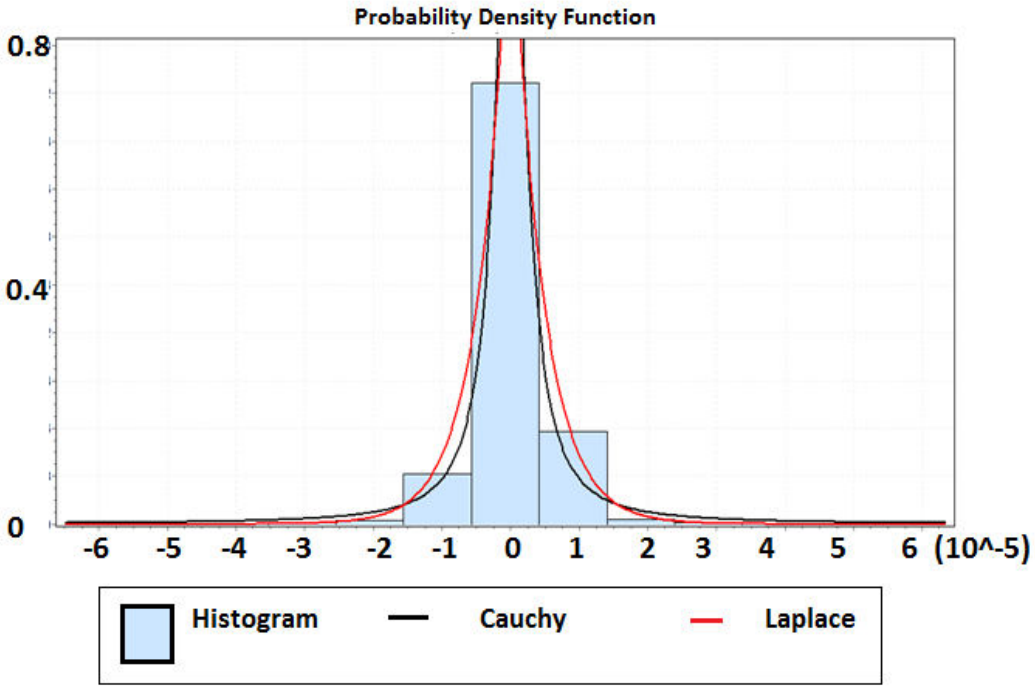


**Figure 2.4 (b) DCT domain histogram of western instrumental Guitar.**

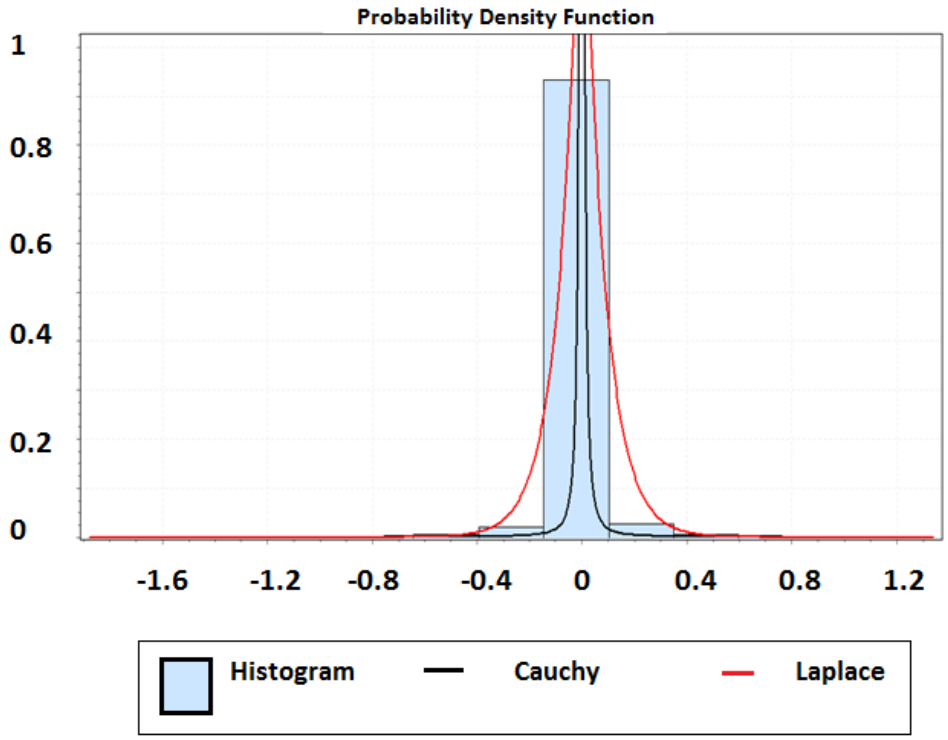


**Figure 2.4 (c) DCT domain histogram of East Indian Folk (Punjabi).**

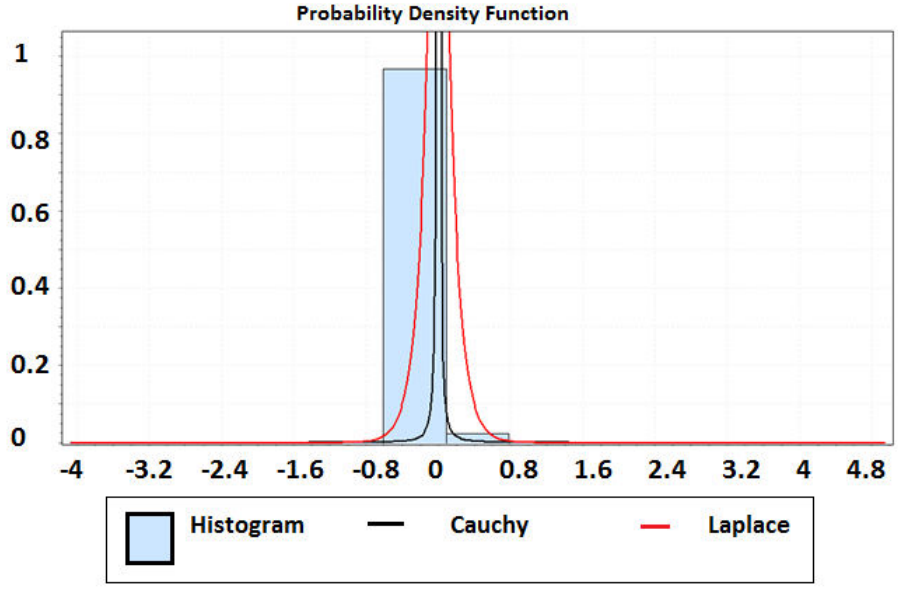


**Figure 2.4 (d) DCT domain histogram of East Indian pop.**

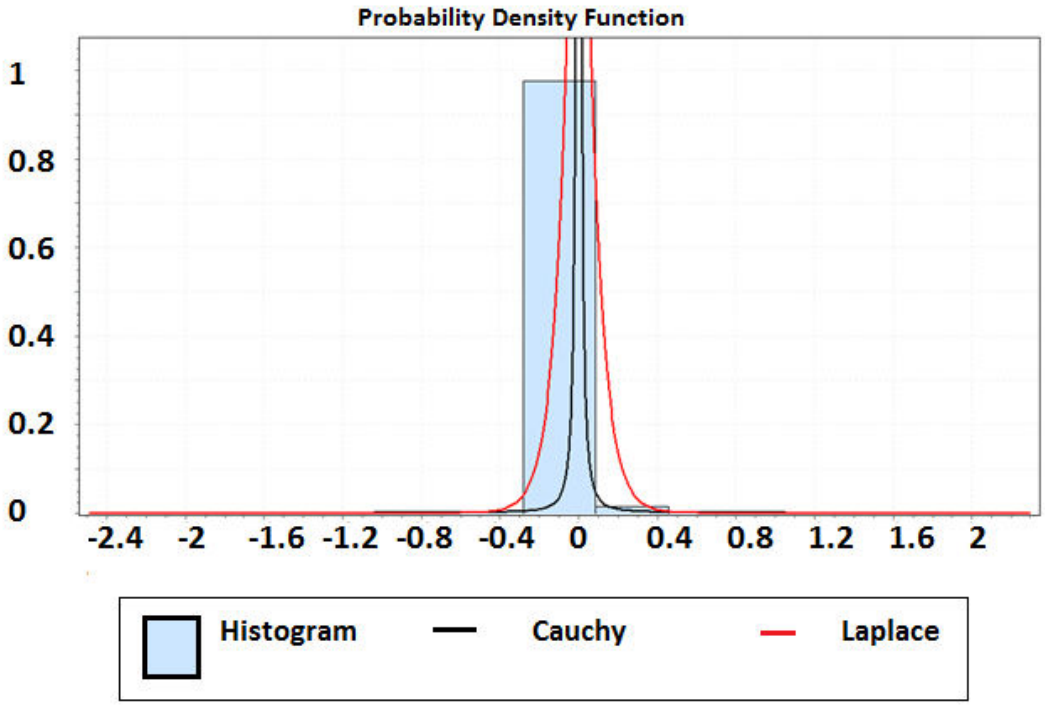


**Figure 2.4 (e) DCT domain histogram of East Indian devotional.**

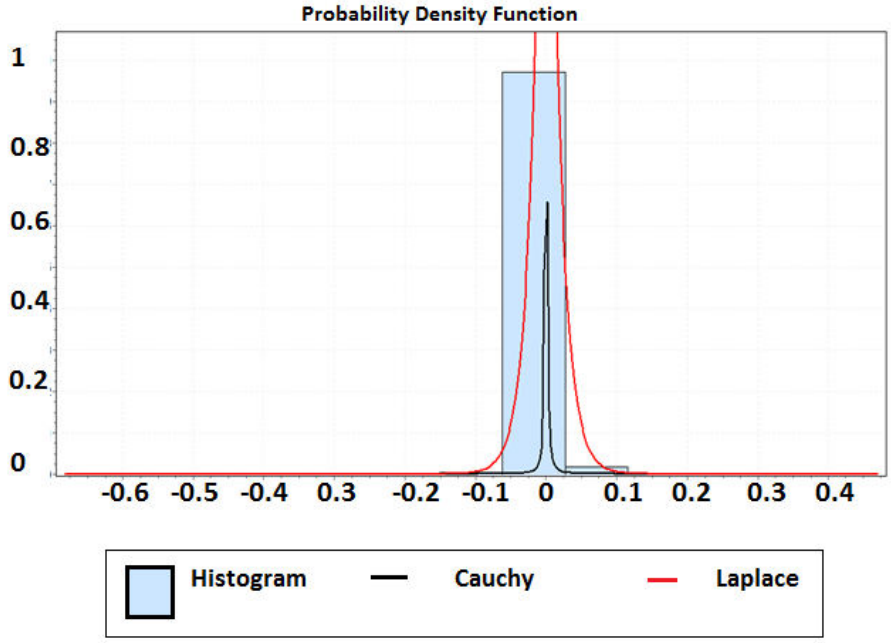


**Figure 2.4(f) DCT domain histogram of Blues**.

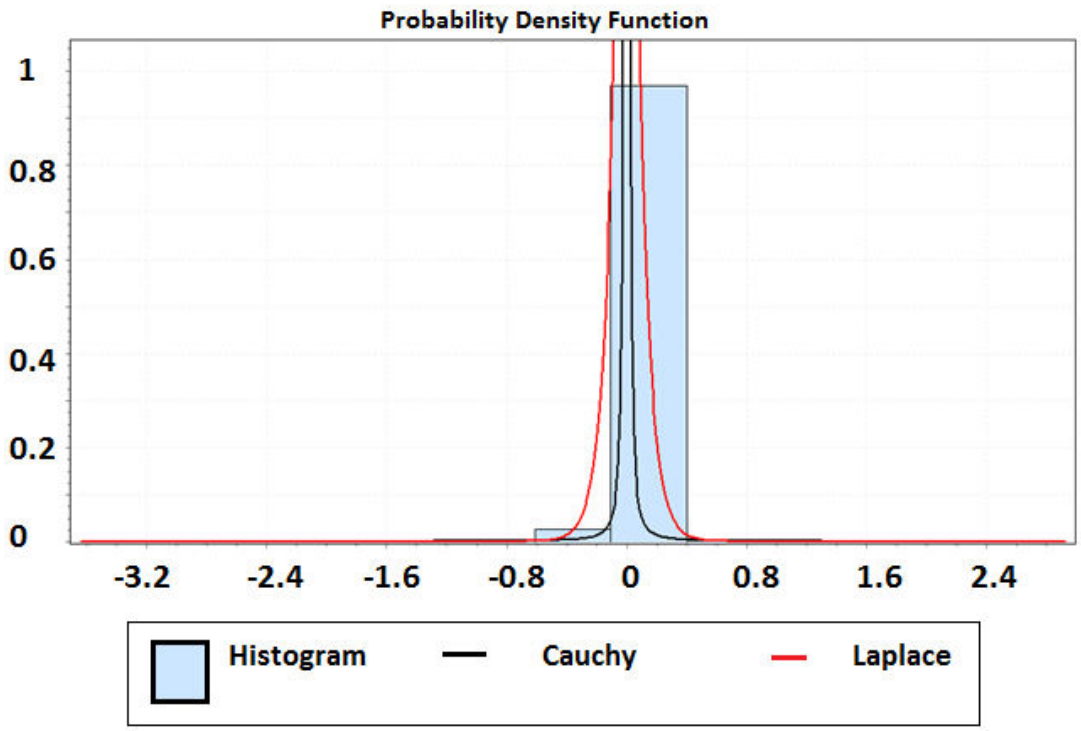


**Figure 2.4(g) DCT domain histogram of Percussion.**

**Table 2.3 Kurtosis and Goodness of fit Tests statistic (DCT domain)**

| Sr no. | Distribution | Kolmogorov Smirnov(test statistic) | Anderson Darling(test statistic) | Chi-Squared(test statistic) |
|---|---|---|---|---|
| Jazz | | | | |
| 1 | Cauchy | 0.03078 | 10.091 | 197.38 |
| 2 | Laplace | 0.14317 | 212.49 | 1584.7 |
| Western Instrumental (Guitar) | | | | |
| 1 | Cauchy | 0.05767 | 39.394 | 693.83 |
| 2 | Laplace | 0.04339 | 21.673 | 227.61 |
| East Indian Folk (Punjabi) | | | | |
| 1 | Cauchy | 0.13873 | 188.08 | 5205.4 |

| | | | | |
|---|---|---|---|---|
| 2 | Laplace | 0.29592 | 809.32 | 20611.0 |
| East Indian Pop | | | | |
| 1 | Cauchy | 0.0628 | 34.978 | 1009.6 |
| 2 | Laplace | 0.34815 | 1095.5 | 27722.0 |
| Blues | | | | |
| 1 | Cauchy | 0.1247 | 281.51 | 4428.1 |
| 2 | Laplace | 0.35113 | 1056.0 | 30223.0 |
| East Indian Devotional (Classical) | | | | |
| 1 | Cauchy | 0.17895 | 248.52 | 9555.2 |
| 2 | Laplace | 0.24873 | 751.87 | 13238.0 |
| Percussion | | | | |
| 1 | Cauchy | 0.08416 | 116.75 | 1469.6 |
| 2 | Laplace | 0.27062 | 687.23 | 17779.0 |

### 2.4 Estimation In The Wavelet Domain

The wavelet theory has found many application in signal processing [23] by offering highly efficient algorithms. Wavelet analysis has many advantages over traditional Fourier transform, it can operate on both time and scale aspects with the 'scale' aspect containing frequency information and thus time frequency localization is achieved [24-26].

There is a direct correspondence between wavelet scale and frequency. The mother wavelet is compared with the signal to be analyzed at different scale and the degree of similarity at a particular scale and time gives us the wavelet coefficient at that particular scale and time. The larger the scale, the more stretched is the wavelet and hence the longer the portion of the signal with which it is being compared. This gives a set of wavelet coefficients which measure the coarser signal features which in turn correspond to low frequency. Thus the part of signal having lower frequency would give higher coefficients at higher scale and vice-versa. However, proper scale to frequency transformation is required to yield a near perfect time frequency analysis. In this work,

the music signals have been subjected to continuous wavelet transform. Continuous wavelet transform was chosen since it is more ideal candidate for time-series analysis due to its more fine grained resolution.

The continuous wavelet transform (CWT) of a continuous, square-integrable function x(t) at a scale a>0 and shift b ,is given by

$$X_{\omega}(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t)\psi^{*}\left(\frac{t-b}{a}\right)dt \qquad (2.5)$$

Whereas, in the Fourier transform the signal values are weighted with an exponential argument, in the wavelet transform, signal values are weighted by wavelet function.

The music signals were subjected to CWT at different scales and the set of coefficients generated at each scale have been analyzed to estimate their distributions. The results obtained have been shown in Figs. 2.5(a)-(b)

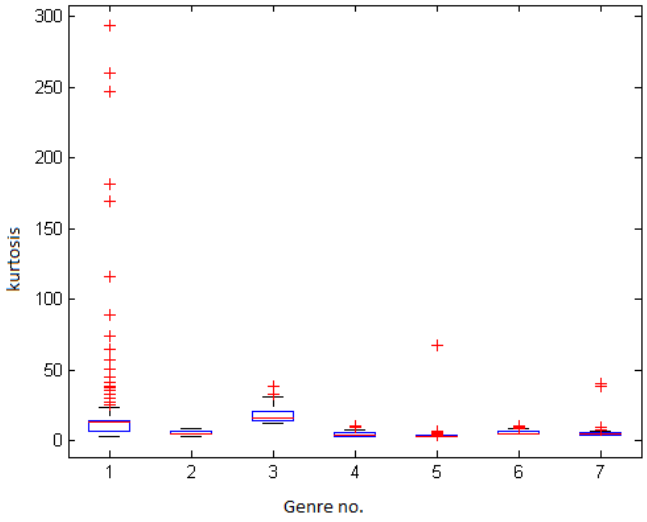


**Figure 2.5(a) Raw coefficients.**

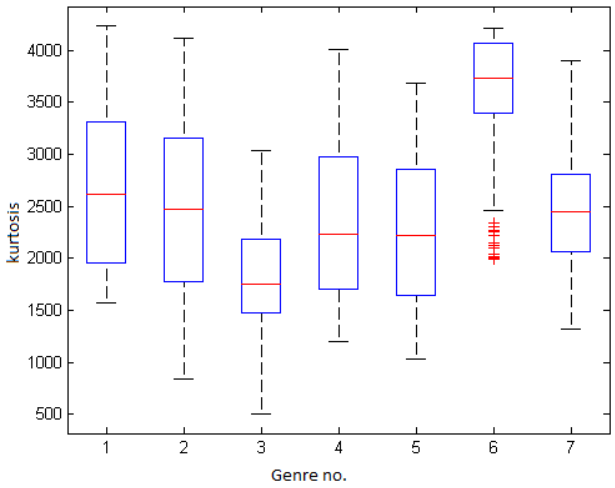


**Figure 2.5(b) MS PCA selected coefficients.**

The author has used db 4 wavelet for decomposition as it is the most widely used. Coefficients generated at each scale have Laplacian nature. However, skew and kurtosis exhibit a lot of variation. It is also evident from Fig 2.5(a)-Fig 2.5(b) that the skew at higher scale deviates from the values obtained using samples in the time domain. This can be attributed to "oversolving" the signal using too many scales for analysis.

To get a clearer picture the author has subjected the wavelet coefficients to Multiscale-Principal-Component-Analysis (MSPCA) which selects the eigenvectors of the covariance matrix. Signals are projected on the corresponding wavelet basis function and their lowest frequency content is represented on a set of scaling functions. Let us represent projection on the scaling function by convolution with a filter H, and projection on a wavelet by convolution with a filter G, the coefficients at a different scale may be obtained as

$$a_m = Ha_{m-1} \quad , d_m = Ga_{m-1} \tag{2.6}$$

Where $d_m$ is the vector of wavelet coefficients at a scale m, and $a_m$ is the vector of scaling function coefficients.

At $a_0$ , original signal vector x can be found.

Thus,

$$a_m = H_m X \quad , \quad d_m = G_m X \tag{2.7}$$

Where $H_m$ is obtained by applying the *H* filter m times,and $G_m$ is obtained by applying the $H$ filter (m-1) times and the *G* filter once. The original data may be reconstructed at all scales,$a_L$.

The principal-component are selected from the pool of Wavelet coefficient as follows :-
If X is the original $n \times p$ data matrix which is represented as

$$X = SL_1^T \tag{2.8}$$

Where L are principal component loadings and S are principal component scores , and n and p are the number of measurements and variables respectively.
Let the wavelet transform of $WX$ convert the matrix $X$ into matrix $WX$ of same dimension.
The PCA of wavelet transformed matrix WX is given as

$$WX = (WS)L^T \tag{2.9}$$

The covariance of the wavelet transformed matrix can now be written in terms of contribution at multiple scales as

$$(WX)^T(WX) =$$
$$(H_LX)^T(H_LX) + (G_LX)^T(G_LX) + .. + (G_mX)^T(G_mX) + .. + (G_1X)^T(G_1X) \tag{2.10}$$

The resulting scores at each scale are decorrelated due to wavelet decomposition.The final covariance matrix can be computed by evaluating the covariance matrix at a coarser scale in a scale recursive manner by incorporating the covariance matrix of the wavelet coefficients at the same scale as

$$(H_{m-1}X)^T(H_{m-1}X) = (H_mX)^T(H_mX) + \gamma(G_mX)^T(G_mX) \tag{2.11}$$

$$\gamma = \begin{cases} 1 \text{ if the PCA at scale m has significant events} \\ 0 \text{ otherwise} \end{cases} \tag{2.12}$$

## 2.5 Analytical Justification For Laplacian Assumption

Although some genres have given good results in the CWT domain with Generalized Gamma distribution, and in the time and DCT domains with Cauchy distribution, the author still believes the Laplacian to be the best estimate.

This assumption is based upon the observation that the variance of DCT coefficients and MSPCA selected wavelet coefficient follows a nearly half exponential distribution as evident from the histogram plots of DCT and CWT coefficients depicted in Figures 2.6(a)-2.6(b).

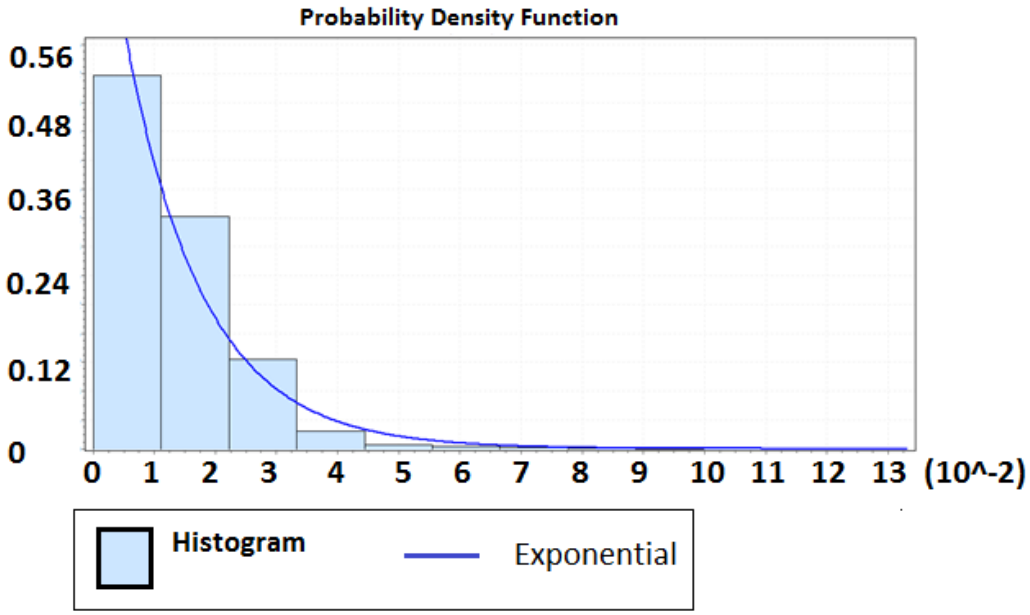


**Figure 2.6(a) Histogram of varience of divergence of jazz.**

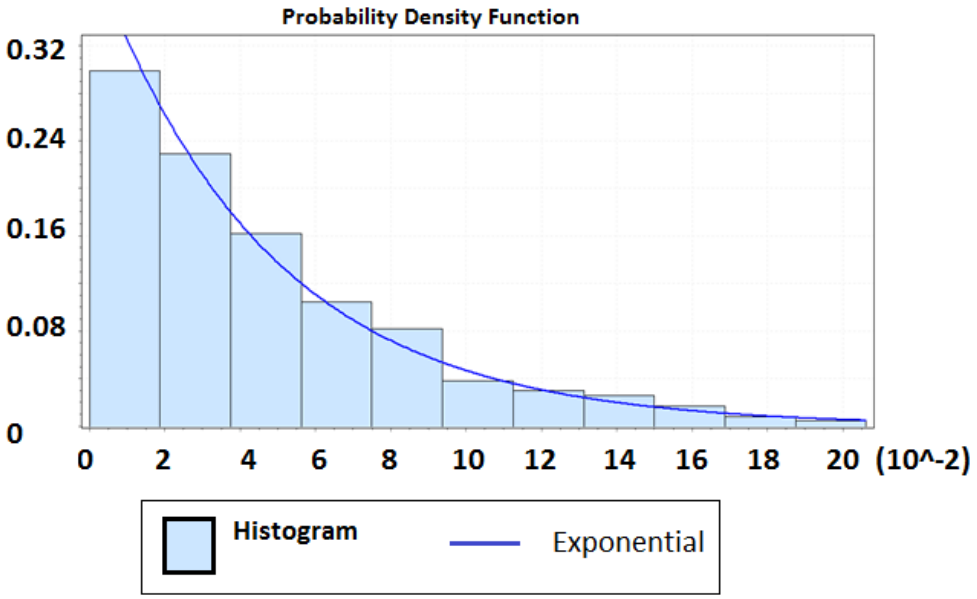


**Figure 2.6(b) Histogram of varience of divergence of instrumental.**

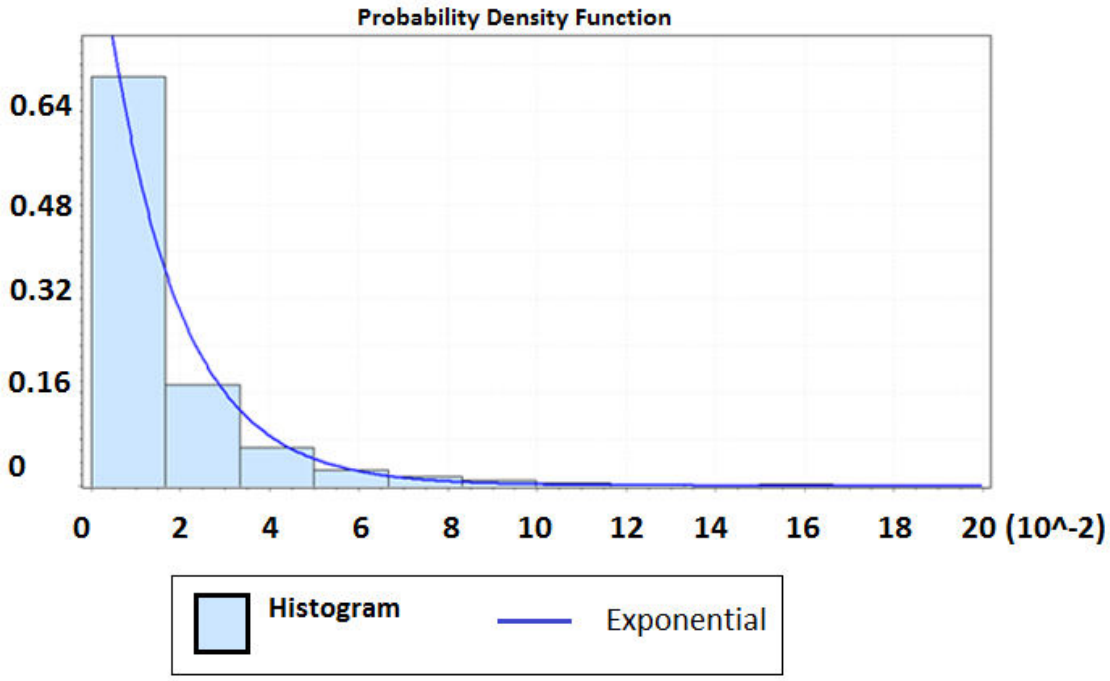


**Figure 2.7(a) Histogram of the Cwt of pop music.**

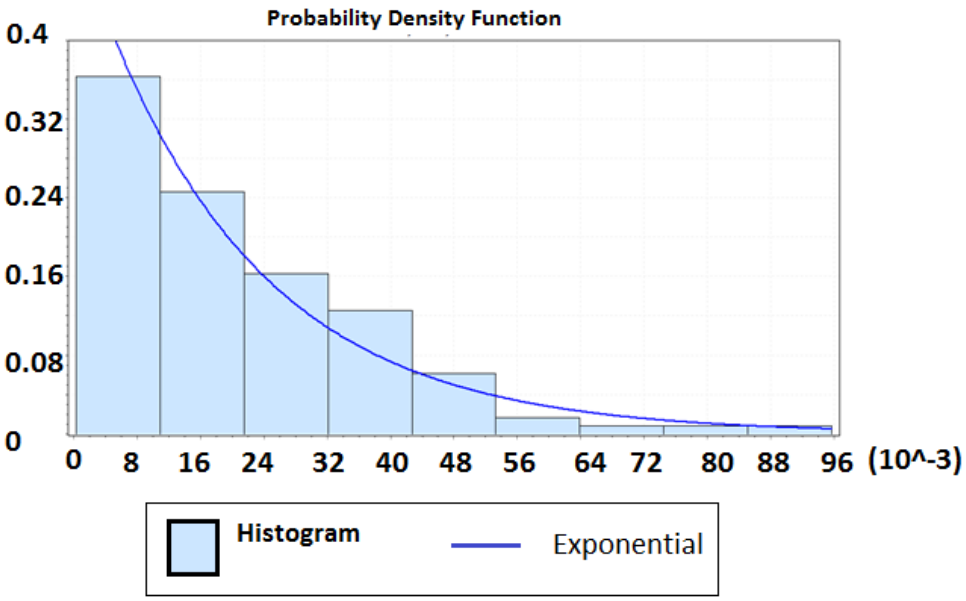


**Figure 2.7(b) Histogram of the Cwt of devotional music.**

Now, if '$s'$ denotes a 10ms sample of a particular recording $'S'$, it can be viewed as an identically distributed random variable. Therefore, a recording which is described by the sum of the individual samples in the decorrelated domains can be approximately Gaussian.

Let, $P(\frac{S}{\sigma^2})$ be approximately Gaussian according to central limit theorem. Where, $\sigma^2$= variance of the underlying distribution.

Thus, $$P\left(\frac{S}{\sigma^2}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{-\frac{S^2}{2\sigma^2}\right\} \quad (2.13)$$

The conditional probability can be given as

$$P(S) = \int_0^\infty P\left(\frac{S}{\sigma^2}\right) P(\sigma^2) d(\sigma^2) \quad (2.14)$$

Now, according to Figs. the variance of DCT and CWT coefficients have exponential nature.

$$\therefore P(\sigma^2) = \lambda \exp\{-\lambda\sigma^2\} \tag{2.15}$$

Using (2.15) in (2.14) and (2.13),

$$P(S) = \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{-\frac{S^2}{2\sigma^2}\right\} \lambda \exp\{-\lambda\sigma^2\}\, d(\sigma^2) \tag{2.16}$$

Rearranging,

$$P(S) = \sqrt{\frac{2}{\pi}}\lambda \int_0^\infty exp\left\{-\lambda\sigma^2 - \left(\frac{S^2}{2}\right)\frac{1}{\sigma^2}\right\} d(\sigma^2) \tag{2.17}$$

However,

$$\int_0^\infty exp\left\{-ax^2 - \frac{b}{x^2}\right\} dx = \frac{1}{2}\sqrt{\frac{\pi}{a}} \exp\{-2\sqrt{ab}\} \tag{2.18}$$

$$\therefore P(S) = \left(\sqrt{\frac{2}{\pi}}\lambda\right)\left(\frac{1}{2}\sqrt{\frac{\pi}{\lambda}}\right) exp\left\{-2\sqrt{\frac{\lambda S^2}{2}}\right\} = \sqrt{\frac{2\lambda}{2}} exp\{-\sqrt{2\lambda}\,|S|\} \tag{2.19}$$

But Laplacian function is given by:

$$P(x) = \frac{\mu}{2} exp\{-\mu\,|x|\} \tag{2.20}$$

Hence, $P(S)$is Laplacian with $= \sqrt{2\lambda}$ .

**2.6 Conclusion**

This study brings further insight into the nature of musical signals by estimating the distribution of music belonging to different genres. Unlike speech signals, music signals can be classified in a plethora of ways. It is highly desirable to have a precise idea about the statistical nature of music signals classified according to their psychoacoustic

properties. It was observed during the course of this study that Cauchy distribution in the time and DCT domain and Generalized Gamma distribution in the wavelet domain could be an alternative for the Laplacian distribution which has been the most preferred one for modeling ordinary speech signals. Furthermore, the Laplacian nature is more pronounced in monophonic instrumental music as evident from exceptionally high kurtosis obtained in the time domain. Unlike the pdf of speech signals reported earlier, the higher kurtosis of music signals can be attributed to the absence of silent interval in the sample. There is also a clear demarcation between the distribution of vocal and instrumental music samples. Significantly high peakedness of instrumental music points towards a negative correlation between kurtosis and degree of polyphony. It can also be concluded by this study that Laplacian assumption holds good both because of computational elegance offered by it and some drawbacks associated with other distributions. However, there is a lot of scope to introduce some new distributions particularly suited to genre based music signal analysis which could go a long way in building more efficient automatic transcription systems.

# Chapter 3

## Development of audio_analyser(1.0)

MATLAB is a very useful platform for studying discrete time signals. What makes it so useful is that it has specifically designed toolboxes to study various engineering domain problems. The MIR toolbox is one such Matlab based toolbox developed by Olivier lartillot et al [20] to analyse music signals. Based upon this toolbox more easy to use systems can be built.Also Simulink based models may be integrated to form user friendly programs especially for the new user.

### 3.1 Motivation:-

During this study it was deemed important to quickly view audio characteristics and decide which of these characteristics may be used to build classification systems.Also reloading audio features for different classification programs needed to be avoided.As a result a GUI named 'analyser_audio' was developed combining useful features of the MIR toolbox into one user friendly interface.

### 3.2 Overview:-

This chapter explains the GUI developed during this Dissertation. The GUI 'Analyser_audio' basically consists of a main GUI and 6 linked GUIs to perform audio handling with ease. This GUI is basically designed to use audio parameters so as to classify data and view the classification results (figure[3.1] shows the main GUI window).The main GUI window consists of 14 push buttons and 5 list boxes .The subsequent sections explain in detail the rest of the GUIs followed by a simple session. The GUI combines upto a 100 functions to primarily focus upon the following tasks:-

1) Providing an interface to view the time and frequency domain parameters of a segment and/or the complete song.
2) Handle the paramaters of single as well as multiple audio tracks.
3) Speed up the process of classification.
4) To record and denoise audio using wavelet coefficients.

5) To view scatter plots of different genres
6) Generate a Excel sheets which contain features extracted from the test as well as the training folder.

**3.3 Installation Requirements**:-

The GUI requires MATLAB version 7.0 or later and the following toolboxes installed:-

1) MIR toolbox .
2) Signal Processing Toolbox.
3) Wavelet toolbox.
4) Auditory toolbox.

**3.4 Installation:-**

Audio_analyser may be installed by copying the functions directly to the current directory or adding the folder to the current search path by using the set path utility.

**3.5 The Main GUI Window**:-

The main GUI window provides the user with a link to the other subordinate GUIs and also a platform for viewing classification results.

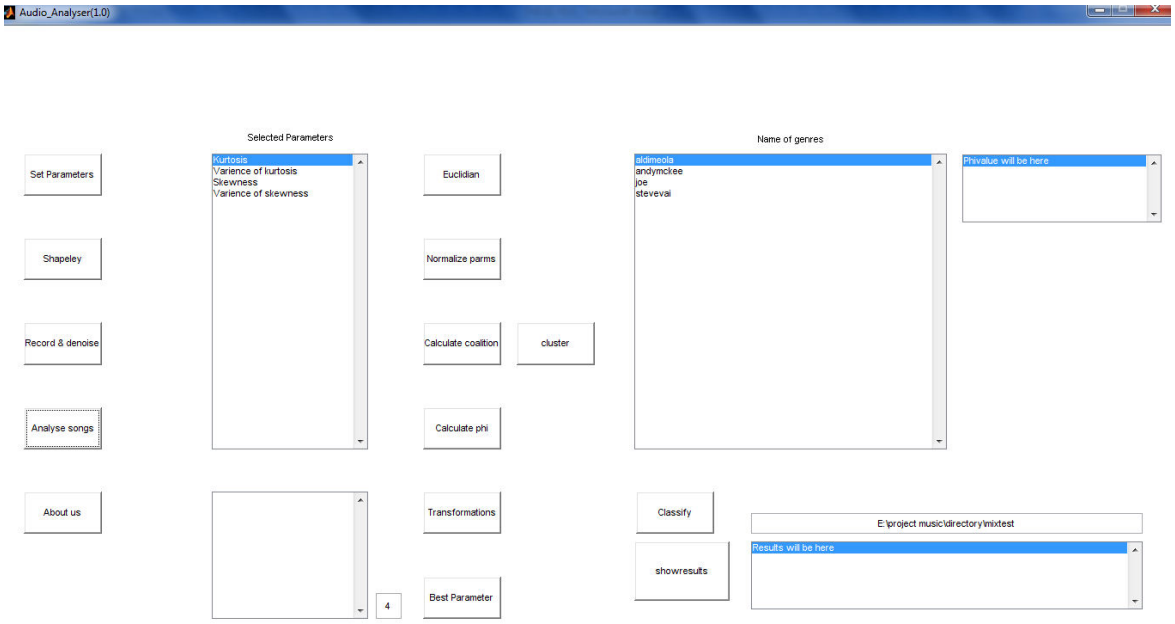


**Figure 3.1 audio_analyser main window.**

### 3.6 Song analyser window:-

The Song_Analyser GUI can be used to selectively listen to any song and view its parameters .This GUI can be accessed using the fourth button in the first column (Analyze Songs).Along with the main window two graphs appear when the song analyser window is opened. Figure (3.3) and (3.4) show the time domain representation and the histogram of a sample song.

The listbox on the left top clearly displays the contents of the current directory and any song file (wav or mp3 format) may be directly loaded into the GUI simply by using shift +left click. The static text on the top of the listbox displays the current directory. The play, pause, resume and stop buttons may be used to play,pause resume or stop the current audio track. Below the list box there are two edit text boxes which the user can use to select the desired time segment of the song. The static text below these values displays the name of the current loaded song .The listbox below them displays the calculated parameters of the song segment selected by taking into account the edit boxes. If the user chooses inappropriate starting or ending values then values for the whole song are computed.

The play /pause buttons may be used in two modes segment mode or the song mode depending upon wheather the user wants to listen to the complete song or only a segment in question. The segment having its duration and starting point according to the edit text values.

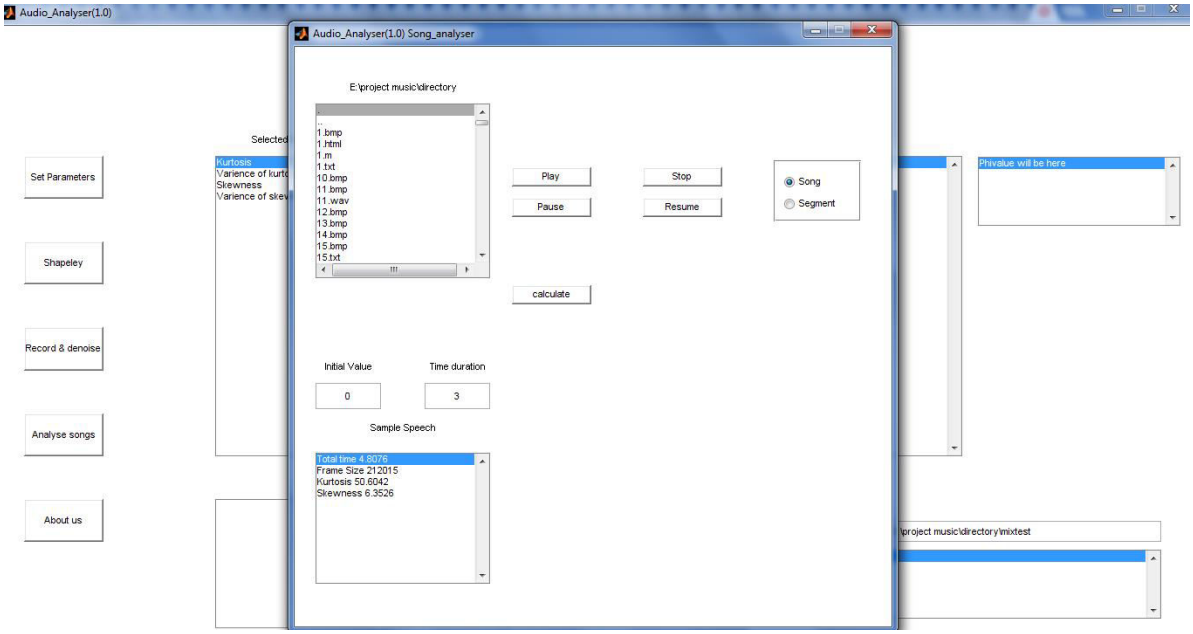


**Figure 3.2 Song analyser window.**

### 3.6.1 Sample Session:-

A sample session may include loading a song by scrolling through the current directory listening to it and focusing on certain aspects of the song which may show different behavior with respect to the whole song .When the calculate button is pressed then all three figure windows will open containing the magnified version of the current song segment, the time domain histogram as well as the DCT domain histogram of the segment as shown in the figure (3.5).

This process may be repeated any no of times for different audio tracks and useful conclusions may then be drawn from range of the parameter values obtained in the second listbox for different types of audio tracks.

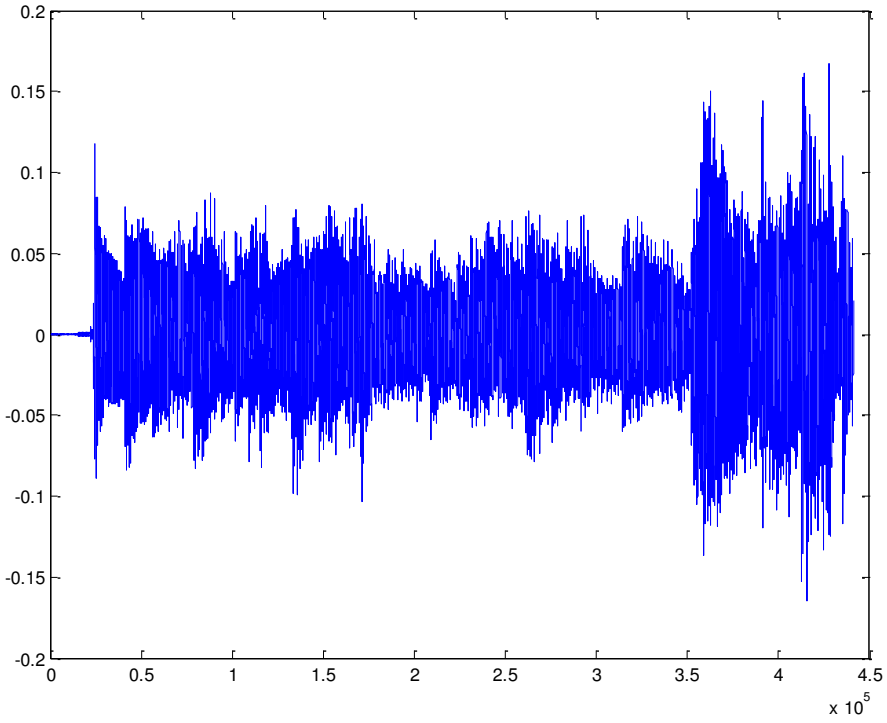


**Figure 3.3 Time domain plot of a sample speech signal.**

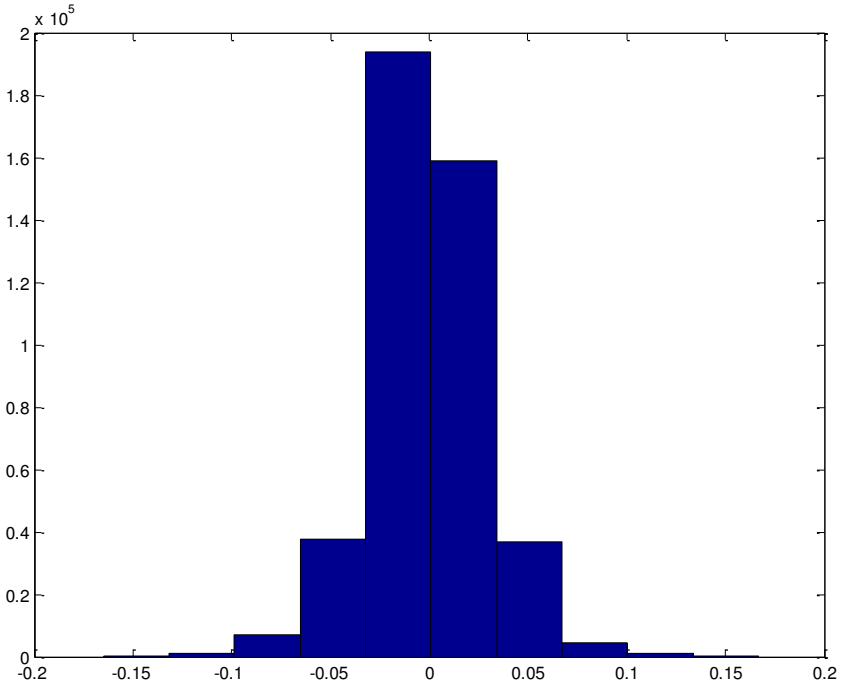


**Figure 3.4 Time domain histogram of a sample speech signal.**

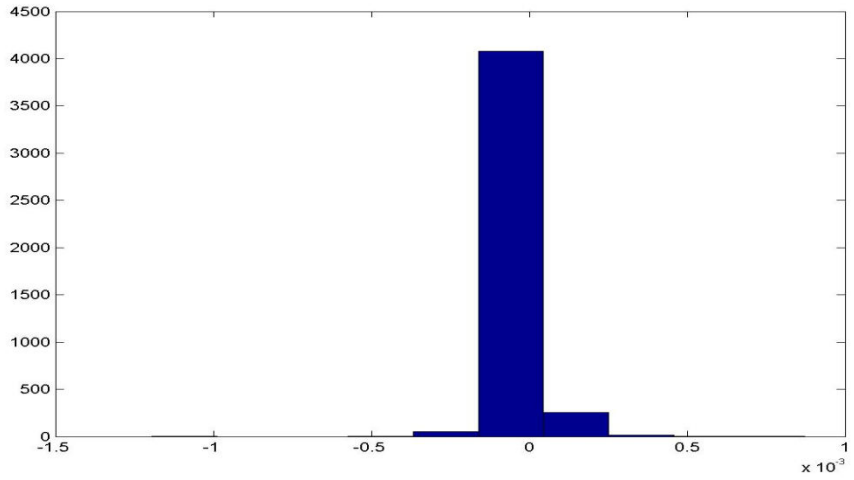


**Figure 3.5 DCT domain histogram of a sample speech signal.**

### 3.7 Song Recorder and Denoiser Window:-

The Song Recorder and Denoiser window may be opened by clicking the third button on the first column in the main analyser_audio window.The primary functions of this GUI is to record live audio and to view graphically or listen to both the orignal file and the denoised file.

The user may choose to record live audio by clicking on the open model button .Clicking on this button will open the 'Simulink_Audio_recorder' model which can be used to record audio this requres an audio input device to function.Figure 3.6 show the simulink model and the Song Recorder and denoiser window.This file may then be denoised using the Denoise button.The author uses wavelets to denoise audio the algorithm used is discussed in chapter 4 .Once the audio is recorded the user will be provided with additional options to save both the orignal file and the denoised file.

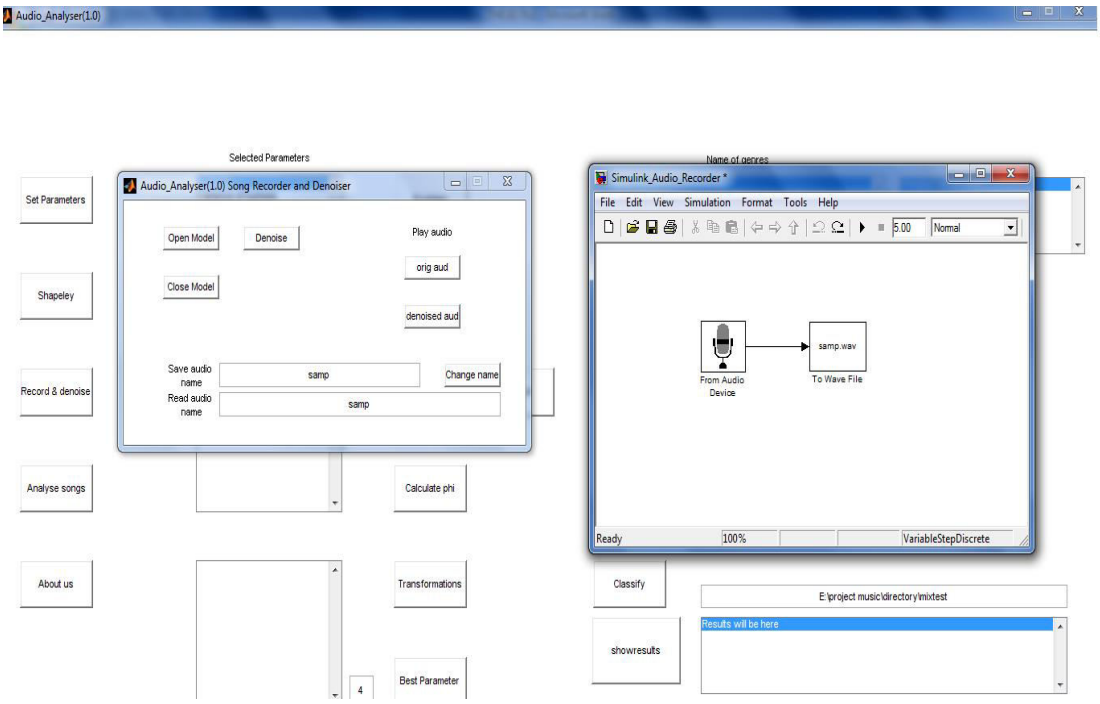


**Figure 3.6 The simulink model and the Song Recorder and denoiser window.**

## 3.8 Transformations window:-

When dealing with classification tasks it is fundamental to view the scatter plots of the features obtained. Keeping this in mind the transformations window was designed.The transformation window can be accessed by clicking on the fifth button on the third Column in the main GUI window.Figure (3.7) shows the

transformations window .

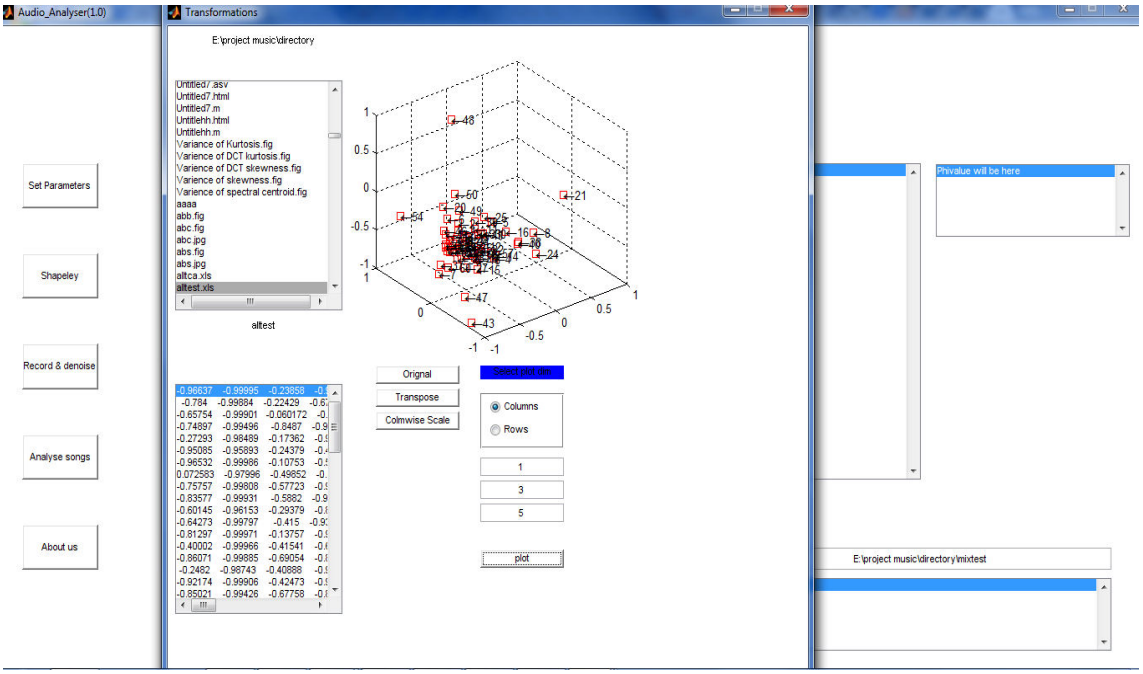


**Figure 3.7 The Transformations window.**

This GUI can basically access any excel file simply by shift clicking on the desired file in the listbox menu. The three buttons may be used to transpose, scale or return to original. The user may choose to plot rows or columns. The user may choose to plot three or two rows or columns by using the edit text buttons .To plot only two dimension enter 0 into the any one of the edit texts and click plot.The results of each plotting session are automatically saved in the current directory in a folder by the name of plots.

## 3.9 Set Parameters Window:-

The user may want to choose different feature vectors for the classification tasks and this can be done by accessing the set parameters window.The set parameters window can be accessed by clicking on the first button on the first column in the main GUI window.

The user may choose to add or remove any of the features from a list box.The feature set is described in chapter 4.Figure (3.7) shows the set parameters window.

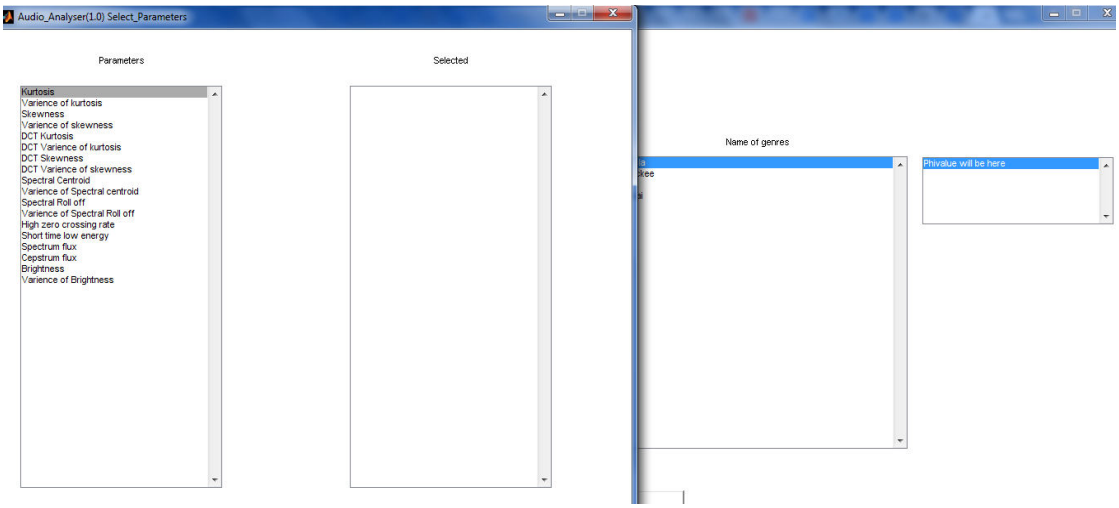


**Figure 3.8 The Select parameters window.**

### 3.10 Sample GUI session:-

The different windows mentioned above may be used individually or in may be used together.Figure (3.8) shows a flowchart for a typical session on the analyser_audio GUI. The steps are sequentially described as under:-

1) Start audio_analyser: -

   The audio_ analyser can be started by typing audio_analyser in the command window. It may also be started from the functions menu. The GUI will change the current directory to the directory where the GUI is installed.

2) Analyse songs:-

   In the beginning the user may want to individually analyse different audio samples and for this the song analyser window can be used. At this point the user may also record audio. Selecting a better feature set is of fundamental importance to the classification task.

3) Copy files and Restart:-

Now the user must arrange the files according to the desired classification (for example genre based or artist based) and copy the desired files into their respective class folders.The GUI is designed to read all classes from a folder named analysis in the current directory so all class folders must be copied into the analysis folder. If restarted the GUI will automatically load the class labels from the analysis folder.

4) Load and generate excel files:-

The features of these audio files may be extracted by clicking on the 'normalize params' button .This sequentially loads all the songs in the analysis folder and extracts features . All these feature vectors are clubbed into one excel file.

One may also choose to skip this step altogether in that case the excel file from the previous session may be used. Most software's are compatible with excel files and therefore these files may be used elsewhere.

5) Select feature vectors:-

The feature vectors to be used may be selected from the Select parameters window .Once the window is closed the main audio_analyser window is automatically updated. In case
the user does not select the feature vectors default values are loaded.

The feature vectors are described in chapter 4. The user may chose up to 18 of these the features may themselves be changed by easily modifying the normalyser function in the GUI code.

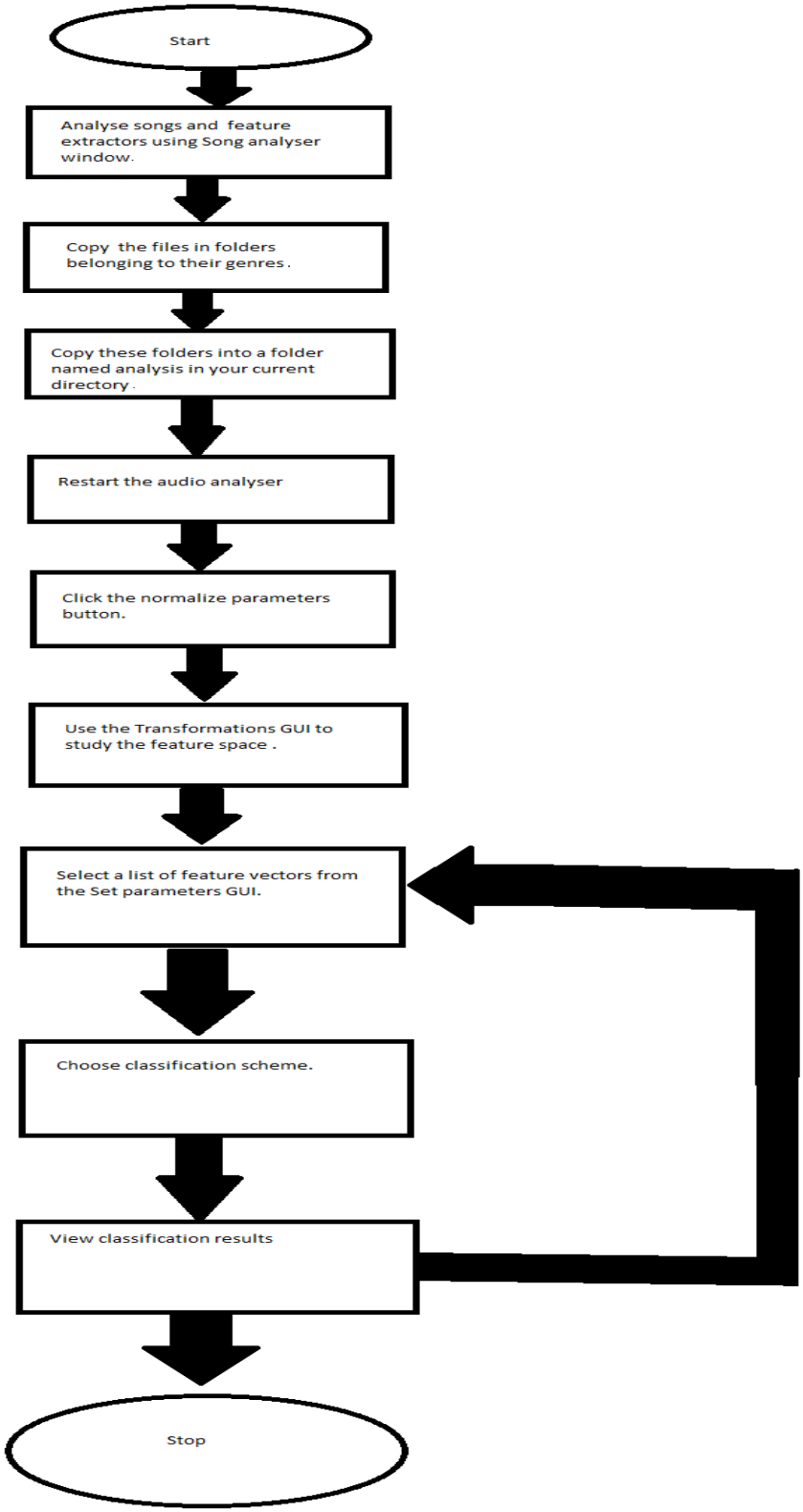


**Figure 3.9 Flowchart for a typical session on the analyser_audio GUI**

6) Choose classification Scheme:-

At this point a classification scheme may be chosen to classify the given data set the audio_analyser GUI provides Euclidean distance based and Shapeley value based classifiers.
The user may want to use ANN based classifiers and therefore may use neuralnetwork toolbox.

7) View Classification results:-

The classifier may then be used to test the test set and the results can be viewed in the form of a pie chart. The confusion matrix for the given classification scheme is updated on the main GUI window.

8) Repeat session:-

The user may choose to repeat the given session with different feature vectors or with different classes which can be done easily. The time taken to reload these feature vectors is avoided in case same classes are considered.

### 3.11 Conclusion and Future scope:-

Version 1.1 of audio_analyser was implemented on MATLAB. The author proposes to integrate the following features in the subsequent versions:-

1) Additional data descriptors for audio classification.
2) Add additional classifiers .
3) Provide better audio denoising with a choice for audio denoising algorithms.
4) Improve transformations by providing a variety of figure import formats.

5) Support multiple audio formats.

# Chapter 4

# Genre Classification with Wavelet Denoising as a Preprocessing Step

## 4.1 INTRODUCTION

Different Cultures from all over the world have played , written and studied music in a variety of ways. The term 'Musical Genre' has often been used to describe these varied musical traditions. Music is increasingly being shared over the internet and so the need for organizing such a large amount of data is apparent[1]. It's imperative to improve classification rates between certain genres such as Metal and Rock . This is evident from the in the popular paper presented by Tzanetakis et al [4]. In this work the author proposes a neural network based classifier which attempts to solve this classification problem. The author also investigates the use of wavelet transforms as a preprocessing step. Wavelets can be used in a variety of ways for example Duraswamy et al[23] proposed an audio denoising technique with the use Of Biorthogonal Wavelet transform,Johnsen et al[24] demonstrated the use of Bionic Wavelet Transform to Enhance the quality of speech signal. Bahoura et al.[25] proposed a new speech enhancement method based on time and scale adaptation of wavelet thresholds. This study is organized as follows Section 4.2 describes the data set followed by Section 4.3 describing the feature vector used to describe the audio track .Then Section 4.4 describes the use of wavelet transform for denoising and in section 4.5 the classifier used is described .Finally Section 4.6 compares the results obtained by raw and the denoised data. We conclude in Section 4.7 with directions to further this study.

## 4.2 Data Set

The GTZAN dataset [http://marsyas.info/download/data_sets] is used in the present study.This data set has been used for studies in this context [26-31]. For both Genres 100 representative tracks were chosen. The tracks include excerpts from radio, compact disks and MP3 compression audio files. These files were stored as 22050Hz 16 bit mono audio tracks.

### 4.3 Features

Feature selection is very important to achieve good classification accuracy. The author uses an analysis window of length 50 ms (non- overlapping) over a texture window Of 10 seconds. The features are:-

1) Spectral Centroid[4]:-The spectral centroid is defined as the center of gravity of the magnitude spectrum of the STFT. It is a measure of spectral shape and higher centroid values corresponding to "brighter" textures with more high frequencies.

$$C_t = (\sum_{n=1}^{N} M_t[n] * n) / \sum_{n=1}^{N} M_t[n] \quad (4.1)$$

Where $M_t$[n] is the magnitude of fourier transform at frame t and frequency bin n.

2) Spectral Roll off[4]:- The spectral roll off is defined as the frequency $R_t$below which 85% of the magnitude distribution is concentrated. It is another measure of Spectral shape.

$$\sum_{n=1}^{R_t} M_t[n] = 0.85 * \sum_{n=1}^{N} M_t[n] \quad (4.2)$$

3) Spectral Flux [4]:-The spectral flux is defined as the Squared difference between the normalized magnitudes of successive spectral distributions.It is a measure of the amount of local Spectral change.

$$F_t = \sum_{n=1}^{N} (N_t[n] - N_{t-1}[n])^2 \quad (4.3)$$

4) High zero crossing rate [4]:-It simply indicates the number of times a signal changes sign.

$$Z_t = \frac{1}{2} * \sum_{n=1}^{N} |sign(x[n]) - sign(x[n-1]) \quad (4.4)$$

5) Brightness [4]:- The percentage of energy concentrated above a fixed frequency value(1500hz in our case).

The kurtosis and skew are calculated in both the time and DCT domain. Then calculate the above features over the analysis window and take their mean and varience over the texture Window forming a 15 dimentional feature vector Figure 1 shows the 3D Scatter plot of this feature vector

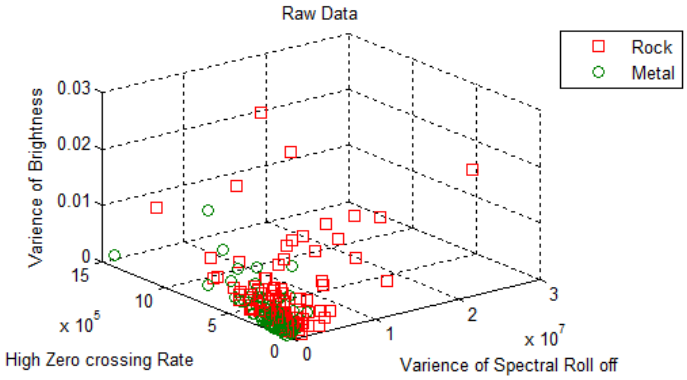


**Figure 4.1 3D scatter plot of Raw audio**

## 4.4 Denoising using Wavelets

Wavelets can be characterized by scale and position. Its advantage over traditional signal processing transforms lies in the fact that the size of the wavelet can vary giving it the ability to process both frequency and time domain data.

It has been experimentally shown [32] that among Symlets 2 to 8, Daubechies 2 to 10 and Coiflet 1 to 5, Coiflet 5, Daubechies 9 and 10 are best suited for denoising audio signals. For the present analysis we use Coiflet 5 wavelet with level 10.We obtain the noise threshold by means of the penalization method suggested by Birge-Massart.
Figure 2 shows original audio followed by figure 3 showing Wavelet denoised audio.

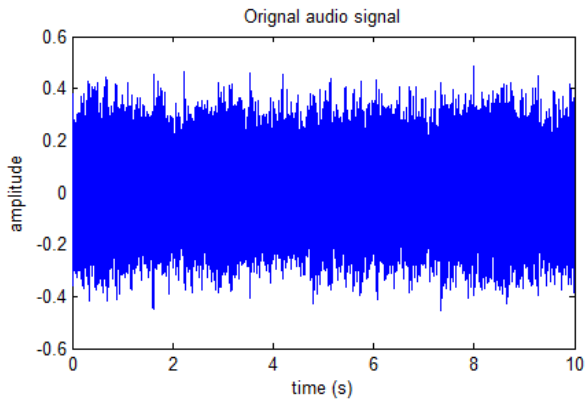


**Figure 4.2 Time domain plot of raw audio signal.**

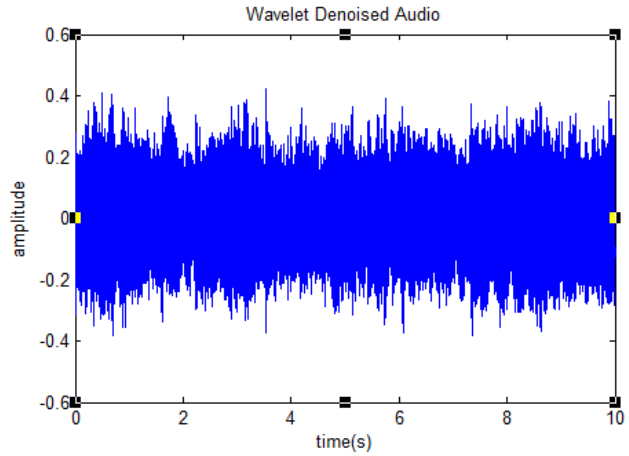


**Figure 4.3 Time domain plot of wavelet denoised audio.**

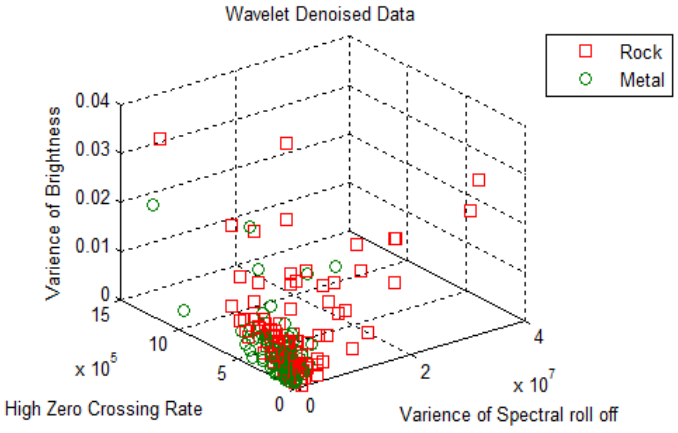


**Figure 4.4 Scatter plot of wavelet denoised audio.**

Figure 4 shows the 3D scatter plot of the feature vector extracted from the denoised audio tracks.

### 4.5 Description of the Classifier

The author uses a two-layer feed-forward network with back propagation, with the no of neurons in the hidden layer varying from 1 to 10

### 4.6 Classification results

Following figures show the boxplot of the training data of both the raw and denoised

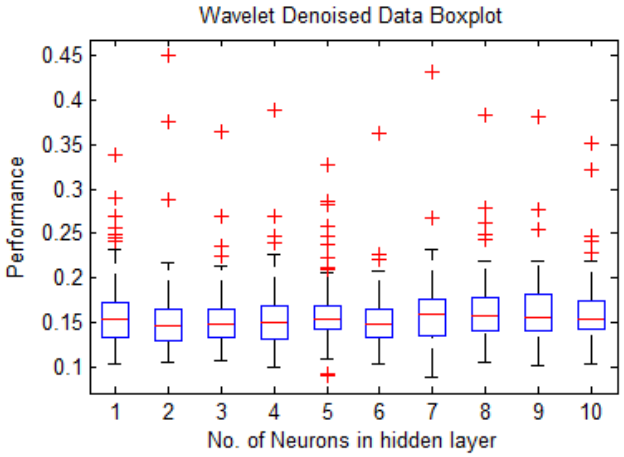


**Figure 4.5 Boxplot of Wavelet denoised data.**

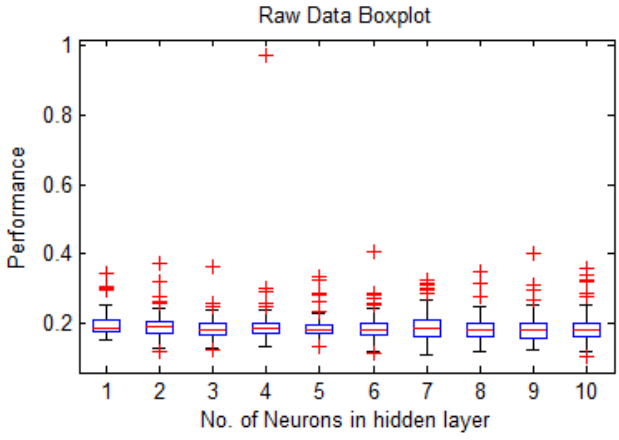


**Figure 4.6 Boxplot of raw data.**

audio. Out of a total of 200 test song samples (100 belonging to each) The authors obtained the results as shown in the confusion matrix in table each column shows the no of correctly identified samples from each class. Classification accuracy of upto 78% was achieved on GTZAN data set.

**Table 4.1 Confusion Matrix of Raw and Wavelet denoised data**

| | Raw | | Wavelet Denoised | |
|---|---|---|---|---|
| No of neurons In Hidden layer | Metal (100samples) | Rock (100samples) | Metal (100samples) | Rock (100 samples) |
| 1 | 89 | 52 | 62 | 87 |
| 2 | 63 | 86 | 78 | 78 |
| 3 | 79 | 76 | 82 | 71 |
| 4 | 51 | 72 | 88 | 55 |
| 5 | 72 | 77 | 73 | 77 |
| 6 | 84 | 70 | 88 | 50 |
| 7 | 77 | 63 | 76 | 74 |
| 8 | 85 | 50 | 81 | 56 |
| 9 | 80 | 51 | 82 | 50 |
| 10 | 81 | 58 | 85 | 63 |

# Chapter 5

## Conclusions and Future Scope

**Musical** signals being more difficult to analyze than their Speech counterparts, pose unseen computational challenges in a plethora of tasks. However we can approach any given problem in one of the two ways either modeling based or by model free computation.

In this Dissertation the author has endeavored to incorporate both the modeling based and the model free approaches for analysis of musical signals. In the beginning extensive study was undertaken to estimate the nature of the probability distribution of music signals and to provide empirical evidence for the choice of a particular distribution. In the later part of the Dissertation, genre classification was attempted using artificial neural network as a model free approximator.

Based upon the results obtained during the course of the study, following conclusions can be drawn:-

- It was observed during the course of this study that Cauchy distribution in the time and DCT domain and Generalized Gamma distribution in the wavelet domain could be an alternative for the Laplacian distribution which has been the most preferred one for modeling ordinary speech signals.
- The Laplacian nature is more pronounced in monophonic instrumental music as evident from exceptionally high kurtosis obtained in the time domain.
- Different genres exhibit different ranges of skew and kurtosis values.
- Significantly high peakedness of instrumental music points towards a negative correlation between kurtosis and degree of polyphony.
- It can also be concluded by this study that Laplacian assumption holds good both because of computational elegance offered by it and some drawbacks associated with other distributions.
- 78% classification between Metal and Rock genres was obtained using artificial neural networks and wavelet denoising as a pre processing step.

However, there is a lot of scope for further improvement and this work can be extended to

- Introduce some new distributions particularly suited to genre based music signal analysis which could go a long way in building more efficient automatic transcription systems.
- Improve classification accuracy by different classification algorithms.
- Incorporate new features within the proposed GUI.

# PUBLICATIONS

1. Ravi Kumar and Vaibhav Arora , “Probability Distribution Estimation of Music Signal in Time and Frequency domains”, in *Proc. IEEE Sponsored 19th International Conference on Digital Signal Processing, DSP 2014, 21-23 August, Hong Kong Polytechnic University, Hong Kong*.(Accepted)

2. Vaibhav Arora and Ravi Kumar ,“Probability Distribution Estimation of Music Signals in Time, Frequency, and Wavelet Domains,” *IEEE Trans. on Audio Speech and Language Processing*.(Communicated)

# REFRENCES


[1] M. Muller, Daniel P.W., Ellis, A. Klapuri, and G. Richard, "Signal Processing for Music Analysis" *IEEE Journal of Selected Topics in Signal Processing,* vol. 0, no.0, 2011.

[2] B. H. Story, "An Overview of Physiology, Physics, and Modeling of the Sound Source for Vowels" *Acoust. Sci. & Tech.* vol. 23, no. 4, 2002, pp. 195-206.

[3] Y. Shiu, and C.C. Jay Kuo, "Music Beat Tracking via Kalman Filtering and Noisy Measurements Selection, In *Proc. IEEE International Symposium on Circuits and Systems, ISCAS 2008*, 18-21 May 2008, pp. 3250-3253 .

[4] G. Tzanetakis, P. Cook, "Musical genre classification of audio signals," *IEEE Tr. Speech and Audio Processing*, vol10 ,no 5,2002,pp. 293-302.

[5] Miguel Alonso, Bertrand David, Ga¨el Richard, "Tempo and beat estimation of musical signals"*ENST-GET, D´epartement TSI 46, rue Barrault, Paris 5634 cedex 13, France (2004).*

[6] İlker Bayram and Mustafa E. Kamasak, "A Simple Prior for Audio Signals," *IEEETransactions on audio, speech, and language processing,* vol. 21, no. 6, june 2013, pp. 191-201.

[7] G. Bao, Zhongfu Ye, and Xu, Y. Zhou, "A Compressed Sensing Approach to Blind Separation of Speech Mixture Based on a Two-Layer Sparsity Model" *IEEE Trans. Audio, Speech, Lang. Process,* vol. 21,no. 5, May 2013, pp. 899-906.

[8] Joe Cheri Ross, Vinutha, T. P.and Preeti Rao, "Detecting melodic motifs from audio for Hindustani classical music," *Department of Computer Science and Engineering*

*Department of Electrical Engineering Indian Institute of Technology Bombay, Mumbai 400076, India (ISMIR 2012),* pp 193-198.

[9] Gopala K. Koduri, Joan Serr`a, and Xavier Serra, "Characterization of intonation in carnatic music by parametrizing pitch histograms," *(ISMIR 2012) Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain Artificial Intelligence Research Institute (IIIA-CSIC), Bellaterra, Barcelona, Spain respectively* ,pp. 198-204.

[10] Y. Ueda, Y. Uchiyama, T. Nishimoto, N. Ono, and S. Sagayama, " HMM-based approach for automatic chord detection using refined acoustic features," In *Proc. 35nd IEEE Int. Conf. Accost., Speech, Signal Process. (ICASSP)*, Dallas, TX, 2010, pp. 5518–5521.

[11] M. Mauch and S. Dixon, "Approximate note transcription for the improved identification of difficult chords," In *Proc. 11th Int. Soc.Music Inf. Retrieval Conf. (ISMIR)*, *Utrecht, The Netherlands, 2010,* pp. 135–140.

[12] H. Papadopoulos and G. Peeters, "Simultaneous estimation of chord progression and downbeats from an audio file," In *Proc. IEEE Int. Conf. Accost., Speech, Signal Process. (ICASSP)*, 2008, pp. 121–124.

[13] Andre Holzapfel and Yannis Stylianou, "Beat tracking using group delay based onset detection," *Institute of Computer Science, FORTH, Greece, and Multimedia Informatics Lab, Computer Science Department, University of Crete(ISMIR 2008),*pp. 653-658.

[14] M. Goto, "An audio-based real-time beat tracking system for music with or without drum-sounds," *J. New Music Res.*, *vol. 30, no. 2, pp. 159–171, 2001.*

[15] S. Dubnov, "Polyspectral Analysis of Music Timbre", Ph.D. Thesis, Hebrew University, Israel, 1996 .

[16] W.B. Davenport, “A Study of Speech Probability Distribution” Ph.D. Thesis, Research Laboratory of Electronics, MIT, Cambridge, Mass.,USA, August 1950 .

[17] J. W. Shin, J.H. Chang, and N. S. Kim, “Statistical Modeling of Speech Signals Based on Generalized Gamma Distribution”, *IEEE Signal Processing Letters,* vol. 12, no. 3, March 2005, pp. 258-261.

[18] S. Gazor, and W. Zhang, “Speech Probability Distribution” *IEEE Signal Processing Letters,* vol. 10, no. 7, July 2003, pp. 204-207.

[19] Olivier Lartillot, “MIR toolbox 1.4.1”, written in MATLAB®, MATHWORKS Inc., USA, 2012.

[20] J. P. LeBlanc and P. L. De Leòn, “Speech separation by kurtosis maximization,”In *Proc. ICASSP*, vol. 2, 1998, pp. 1029–1032.

[21]Abdul Mawla M. A. Najih, Abdul Rahman bin Ramli, V. Prakash, and Syed A.R., “Speech Compression Using Discrete Wavelet Transform”,*In Proceedings IEEE Sponsored Fourth National Conference on Telecommunication Technology Shah Alam, Malaysia*, 2003, pp. 1-4.

[22] Aparna R. Gurijala, and J.R. Deller, Jr., “Speech Watermarking with Objective Fidelity and Robustness Criteria, In *Proc. 37th Asilomar Conference on Signals, Systems, and Computers, Monterey*, October 2003, pp. 211-215.

[23] Shankar B. J & Duraiswamy K.,“ Wavelet-Based Block Matching Process: An Efficient Audio Denoising Technique,” *European Journal of Scientific Research,* vol.48 ,no.1, pp.16-28, 2010.

[24] Johnson M. T, Yuan X and Ren Y, "Speech Signal Enhancement through Adaptive Wavelet Thresholding," *Speech Communications*, vol. 49, no. 2,2007,pp. 123-133.

[25] Bahoura M & Rouat J, "Wavelet speech enhancement based on time–scale adaptation," *Speech Communication*, vol. 48, no. 12, 2006, pp. 1620-1637.

[26] K. Chang, J. Jang and C. Iliopoulos, "Music genre classification via compressive sampling," In *Proc. 11th International Conference on Music Information Retrieval (ISMIR)*, 2010,pp. 387-392.

[24] Philippe Hamel, Matthew E. P. Davies, Kazuyoshi Yoshii and Masataka Goto, "Transfer Learning In Mir: Sharing Learned Latent Representations for Music Audio Classification and Similarity," In *Proc. International Society for Music Information Retrieval (ISMIR)* ,2013,pp 9-11.

[25] Allesantro L. Koerich, " Improving the Reliability of Music Genre Classification using rejection and verification," In *Proc. International Society for Music Information Retrieval (ISMIR),* , 2013,pp 511-516.

[26] Stephane Dupont and Thiery Ravet, "Improved Audio Classification using a Novel Non-Linear dimentionality reduction ensemble approach," In *Proc. International Society for Music Information Retrieval (ISMIR),* 2013, pp 287-292.

[27] P. Hamel and D. Eck, "Learning Features from Music Audio with Deep Belief Networks," In *Proc. 11th International Society for Music Information Retrieval Conference (ISMIR)*, 2010 ,pp 339-344.

[28] T. Li and G. Tzanetakis, "Factors in automatic musical genre classification," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, New York, 2003.

[29] Sander Dielman,Philemon Brakel and Benjamin Shrauwen, "Audio Based Music Classification with a pretrained Convolutional Network," In *Proc. International Society for Music Information Retrieval (ISMIR)* ,2011,pp. 669-674.

[30] Rudolf Mayer and Andreas Rauber, "Music Genre Classification by ensembles of Audio and Lyrics features," In *Proc. International Society for Music Information Retrieval (ISMIR)* ,2011,pp-675-680.

[31] J. Bergstra, N. Casagrande, D. Erhan, D. Eck and B. Kegl, "Aggregate features and AdaBoost for music classification," *Machine Learning,* June, 2006,pp.473-484.

[32] Adrian E. Villanueva- Luna, Alberto Jaramillo-Nuñez, Daniel Sanchez-Lucero, Carlos M. Ortiz-Lima, J. Gabriel Aguilar-Soto, Aaron Flores-Gil and Manuel May-Alarcon, *Engineering Education and Research Using MATLAB*, InTech,2011.

[33] T. Eltoft, T. Kim, and T.W. Lee, "On the Multivariate Laplace Distribution," *IEEE Signal Processing Letters,* vol. 30, no. 5, May 2006, pp. 300-303.

[34] S. Gazor, and W. Zhang, "Speech Enhancement Employing Laplacian-Gaussian Mixture", *IEEE Trans. Speech and Audio Processing*, vol. 13, no. 5, September 2005, pp. 896-904.

[35] H. Brehm and W. Stammler, "Description and generation of spherically invariant speech-model signal," *Signal Process.*, vol. 12, no. 2, Mar. 1987, pp.119–141.

[36] L. Rabiner and B. H. Juang, *Fundamentals of Speech Recognition*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[37] P. N. Juslin, "Cue utilization in communication of emotion in music performance: relating performance to perception," *Journal of Experimental Psychology: Human Perception and Performance*, vol.26, no.6, 2000,pp.1797-813.

[38] Jixin Liu, and Zheming Lu, "A Multipurpose Audio Watermarking Algorithm Based on Vector Quantization in DCT Domain," *Journal of World Academy of Science, Engineering and Technology*, vol. 31, 2009, pp. 395-400.

[39] Y. Shiu, and C.C. Jay, Kuo, "Musical Beat Tracking via Kalman Filtering and Noisy Measurements Selection," In Proc *IEEE International Symposium on Circuits and Systems,* ISCAS 2008, 18-21 May 2008, pp. 3250-3253.

[40] S. Ewert, M. Muller, V. Konz, D. Mullensiefen, and G.A. Wiggins, "Towards Cross-Version Harmonic Analysis of Music," *IEEE Trans. Multimedia,* vol. 14, no. 3, June 2012 pp. 770-782.

[41] Y. Qi, J.W. Paisley, and Lawrence Caring, "Music Analysis Using Hidden Markov Mixture Model," *IEEE Trans. Signal Processing,* vol. 55, no. 11, November 2007, pp. 5209-5224.

[42] Te-Won Lee,Terrence J. Sejnowski, "Independent Component Analysis for Mixed Sub- Gaussian and Super- Gaussian Sources" *Technische Universitat Berlin AND Howard Hughes Medical Institute Computational Neurobiology Lab Computational Neurobiology Lab The Salk Institute 10010 N. Torrey Pines Road , USA La Jolla, California 92037, USA*.

[43] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Commun. Pure Appl. Math*, 1988, vol. 41, pp. 909–996,