IMPROVEMENT OF AUDIO FEATURE EXTRACTION TECHNIQUES IN TRADITIONAL INDIAN STRING MUSICAL INSTRUMENT

KOHSHELAN A/L SUNDARARAJOO

A thesis submitted in
fulfillment of the requirement for the award of the
Degree of Master of Information Technology

Faculty of Computer Science and Information Technology
Universiti Tun Hussein Onn Malaysia

AUGUST 2015

ABSTRACT

Audio feature extraction is an essential and significant process where audio features are extracted from the audio files whereby the extracted audio features contains relevant audio information. One of the important roles played by the audio features is to improve the classification accuracy. However, the presence of noise in the audio signals which degrades the quality of the extracted features may result in low classification accuracy. Some of the existing audio feature extraction techniques are Mel-Frequency Cepstral Coefficient (MFCC), Linear Predictive Coding (LPC), Local Discriminant Bases (LDB), Zero-Crossing Rate (ZCR) and Perceptual Linear Prediction (PLP). Furthermore, the three frequently used techniques in audio feature extraction are MFCC, LPC and ZCR. Previous research had mentioned the shortcomings of the three techniques on extracting noisy signal. This has been identified in the case of traditional Indian musical instrument where the vibration of string instrument had produced noise in the highest amplitude. Therefore, Zero Forcing Equalizer (ZFE) was proposed to equalize the noise in the highest amplitude. ZFE was integrated with three audio feature extraction techniques, namely MFCC-ZFE, LPC-ZFE and ZCR-ZFE in order to improve the performance of the existing techniques. The results show the best improvement of classification accuracies obtained for the proposed techniques of MFCC-ZFE were 98.2% of classification accuracies with 4.0% of improvement by using kNN. Meanwhile, the combined features of the MFCC-ZFE + LPC-ZFE + ZCR-ZFE have obtained 98.3% of classification accuracies with 9.1% of improvement by using kNN.

ABSTRAK

Pengekstrakan ciri audio adalah satu proses yang penting di mana ciri-ciri audio yang diekstrak dari fail audio mengandungi maklumat yang relevan. Salah satu peranan penting yang dimainkan oleh ciri-ciri audio adalah untuk meningkatkan prestasi pengkelasan audio. Walau bagaimanapun, disebabkan oleh kehadiran bunyi hingar di dalam isyarat audio mengurangkan kualiti ciri-ciri yang diekstrak serta merendahkan prestasi pengkelasan audio. Beberapa teknik pengekstrakan ciri audio yang sedia ada ialah Mel-Frequency Cepstral Coefficient (MFCC), Linear Predictive Coding (LPC), Local Discriminant Bases (LDB), Zero-Crossing Rate (ZCR) dan Perceptual Linear Prediction (PLP), tetapi tiga teknik yang sering digunakan ialah MFCC, LPC dan ZCR. Kajian sebelum ini telah mengenal pasti kelemahan dalam tiga teknik tersebut yang disebabkan oleh getaran daripada alat muzik tradisional India khususnya alat muzik bertali yang menghasilkan bunyi hingar dalam amplitud tertinggi. Oleh itu, Zero Forcing Equalizer (ZFE) dicadangkan untuk menyelaraskan bunyi hingar dalam amplitud tertinggi. Dalam kajian ini, ZFE diintegrasikan dengan tiga teknik pengekstrakan ciri audio, iaitu MFCC-ZFE, LPC-ZFE dan ZCR-ZFE untuk meningkatkan prestasi teknik yang sedia ada. Hasil penelitian menunjukkan peningkatan ketepatan pengkelasan audio yang terbaik diperolehi oleh MFCC-ZFE ialah 98.2% pengkelasan audio dengan 4.0% peningkatan ketepatan pengkelasan audio menggunakan kNN. Selain itu, penggabungan ciri-ciri audio daripada MFCC-ZFE + LPC-ZFE + ZCR-ZFE memperoleh 98.3% pengkelasan audio dengan 9.1% peningkatan ketepatan pengkelasan audio menggunakan kNN.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

f	-	Frequency
dB	-	Decibel
KHz	-	Kilohertz
MFCC	-	Mel Frequency Cepstral Coefficient
LPC	-	Linear Predictive Coding
ZCR	-	Zero Crossing Rate
kNN	-	k-Nearest Neighbor
BNs	-	Bayesian Network
SVM	-	Support Vector Machine
SMO	-	Sequential Minimal Optimization
NB	-	Naïve Bayes
ISI	-	Inter-Symbol Interference
ZFE	-	Zero Forcing Equalizer
MIMO	-	Multiple Input Multiple Output
MLSE	-	Maximum Likelihood Sequence Estimator
CBMIR	-	Content-based Music Information Retrieval
FFT	-	Fast Fourier Transform
RMSE	-	Root Mean Square Error
MAE	-	Mean Absolute Error

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Background of study

Audio feature extraction is a process to obtain a set of audio features from an audio signal whereby the rich set of audio features can be utilized in the computation aspect such as in determining the average, maximum or frequency values that can be plotted to a spectrogram (Dave, 2013; Bullock, 2008). There are two types of extracted features that are low-level features and high-level features (Lerch, 2012; Furht, 2009). Low-level features represent terms in which humans refer to music such as pitch, tempo, amplitude and others. High-level features are considered to have direct (humanly interpretable) and derived from the low-level features such as genre and style. This research will be focusing on the low-level features since the research will extract the audio features from the high amplitude of the signal. Audio feature extraction can contribute for better audio classification accuracy result depending on the quality of the extracted features (Umapathy, Ghoraani & Krishnan, 2010). However, one of the essential drawbacks in the audio feature extraction process is the presence of disturbance such as noise from the high amplitude signal which is produced from certain instrument that may degrade the quality of the extracted features and lead to low classification accuracy (Stulov & Kartofelev, 2014; Dave, 2013; Wolf & Nadeu, 2008; Subramanian, 2006). There are many techniques for audio feature extraction, such as Mel-Frequency Cepstral Coefficients (MFCC), Local Discriminant Bases (LDB), Linear Predictive Coding (LPC), Perceptual Linear Prediction (PLP) and Zero-Crossing Rate (ZCR) (Dave, 2013; Anusuya & Katti, 2011; Umapathy, Krishnan & Rao, 2007). However, three audio feature extraction techniques are selected which are Mel-Frequency Cepstral Coefficients (MFCC), Linear Predictive Coding (LPC) and Zero-Crossing Rate (ZCR) based on their good performance and frequently used by previous researchers (Chougule & Chavan, 2014; Bormane & Dusane, 2013; Anusuya & Katti, 2011; Kumari, Kumar & Solanki, 2010; Gunasekaran & Revathy, 2008).

MFCC is one of the techniques commonly used in digital signal processing (Weeks, 2010). MFCC has been proven as one of the effective techniques in audio feature extraction (Keronen *et al*., 2011; Furht, 2009). Similar to MFCC, LPC is another technique which offers a powerful, yet simple method to extract the audio information (Elminir, ElSoud & El-Maged, 2012; Sheetal & Raut, 2012). On the other hand, ZCR is useful for musical instruments measurement and endpoint detection (detection of the start and end of unvoiced sounds) (Khan, Bhaiya & Banchhor, 2012). Even though these three techniques have performed well in extracting audio features, they have shown some shortcomings in extracting noisy signal especially from string instrument in the domain of traditional Indian musical instrument (Anusuya & Katti, 2011; Chougule & Chavan, 2014; Bormane & Dusane, 2013).

Traditional Indian musical instrument is one of the oldest musical instruments in the world (The Incredible India Travel, 2011). It has contributed immensely in making Indian music famous around the world. Traditional Indian musical instrument can be categorized into four types such as string instruments, percussion instruments, wind-blown instruments and solid instruments (Gunasekaran & Revathy, 2008). Specifically, in the context of content-based audio extraction and audio classification, these instruments have been used many times to identify the characteristics of a single instrument such as pitch or amplitude among multi-instruments (Gunasekaran & Revathy, 2008). However, it is difficult to identify the characteristics of a particular instrument such as string instrument since each instrument creates a sound which produces music in its own ways. Furthermore, these instruments may contribute to noisy signal due to some disturbance from the indoor, outdoor or instrument itself. Among them, string instruments such as Veena contributes more to noise due to sound produced from the vibration of strings (Dave, 2013; Stulov & Kartofelev, 2014; Subramanian, 2006). In this research, string instrument of traditional Indian musical instrument were used since they contribute to more noise that come from the highest amplitude of the audio signal and lead to low quality of extracted audio features.

Therefore, Zero Forcing Equalizer (ZFE) was proposed in this research to be integrated with three audio feature extraction techniques, namely MFCC-ZFE, LPC-ZFE and ZCR-ZFE to equalize the noise which comes from the highest amplitude. ZFE is a linear receiver (also known as an equalizer) used in communication systems (Kaur & Kansal, 2013). The function of ZFE is to invert the frequency response of the channel. ZFE has an ability to equalize the noise which comes from the highest amplitude of the audio signals (Kaur & Kansal, 2013). ZFE is integrated with audio feature extraction techniques to overcome the shortcoming occur in the existing techniques as well as to improve the quality of extracted features. Furthermore, another linear equalizer known as Minimum Mean Square Error (MMSE) equalizer is also integrated with the three audio feature extraction techniques for performance comparison.

A good audio extraction technique will lead to better accuracy of audio classification. In order to evaluate the performance of the proposed audio feature extraction techniques, five benchmark classifiers were selected. These classifiers were K-Nearest Neighbor (kNN), Bayesian Network (BNs), Support Vector Machine (SVM), C4.5 decision tree and Naïve Bayes (NB) classifiers. The classifiers were selected based on their good performance in audio classification (Kumar, Pandya & Jawahar, 2014; Bello, 2013; Rocha, Panda & Paiva, 2013; Witten, Frank & Hall, 2011; Nettleton, Orriols-Puig & Fornells, 2010; Li, Ogihara & Li, 2003). The classification results are compared to show the performance of the extracted audio features.

1.2 Research Motivation

As the need of reliable information from the audio and music grows, the importance of research on audio feature extraction increases. Audio feature extraction has contributed immensely in various fields such as in data mining involving Content-Based Music Information Retrieval (CBMIR). CBMIR has become a critical research topic and has been given increasing attention in recent years due to the extensive growth in audio and music (Yu *et al*., 2013). CBMIR generally involves analyzing, searching and retrieving music based on audio features of an audio which is normally used to represent songs or music genre. Identifying them would normally involve feature extraction and classification tasks. Theoretically, the greater the features analyzed, the better the classification accuracy can be achieved.

The impact of audio feature extraction in audio classification is huge since the performance of audio classification accuracy can be defined based on the quality of extracted audio features. Furthermore, good quality of extracted audio features may contribute to a better accuracy of audio classification (Umapathy, Ghoraani & Krishnan, 2010). However, the quality of audio features depends upon the behavior of the audio domain. Audio domain, such as traditional India musical instrument is one of the oldest musical instruments in the world, however, there is not much work done in the area of feature extraction compared to Western music (Agarwal, Karnick & Raj, 2013). In the previous research, traditional Indian musical instrument involving string instruments had shown fluctuating behavior in its audio signal during different experimental setups that were identified by using MFCC, LPC and ZCR (Gunasekaran & Revathy, 2008; Chougule & Chavan, 2014). This fluctuating behavior was identified due to the vibration of string instruments which produced unwanted sound (noise) in highest amplitude (Dave, 2013; Stulov & Kartofelev, 2014; Subramanian, 2006). Wolf and Nadeu (2008) also said that MFCC performance degrades severely when the extracted features contain noise. Moreover, Anusuya and Katti (2011) mentioned that if the audio signal used is noisy, the extracted features from MFCC, LPC and ZCR lead to lower classification accuracy.

Based on the previous researches, their findings showed that the existing audio feature extraction techniques are unable to produce high quality audio features due to the presence of noise in the highest amplitude of the audio signal (Stulov & Kartofelev, 2014; Dave, 2013; Wolf & Nadeu, 2008; Subramanian, 2006). This shortcoming have inspired the use of ZFE to equalize the noise in the highest amplitude (Kaur & Kansal, 2013). Therefore, in this research, ZFE was integrated with MFCC, LPC and ZCR to overcome the drawback of audio feature extraction techniques in extracting noisy audio signal in highest amplitude. This research perceived that improvement can be done to the audio feature extraction techniques to overcome the drawback and lead to a better quality of extracted audio features.

1.3 Objectives

This study embarks on the following objectives:

i. To propose three (3) improved techniques of Mel-Frequency Cepstral Coefficient (MFCC), Linear Predictive Coding (LPC) and Zero-Crossing Rate (ZCR) by integrating them with Zero Forcing Equalizer (ZFE);
ii. To implement the proposed techniques in (i) in traditional Indian string musical instrument; and
iii. To evaluate the performance of the proposed audio feature extraction techniques based on audio classification accuracy by using five classifiers which are k-Nearest Neighbor (kNN), Bayesian Network (BNs), Support Vector Machine (SVM), C4.5 decision tree and Naïve Bayes (NB).

1.4 Scope of Research

The scope of research is divided into four parts which are audio acquisition, audio pre-processing, audio extraction and audio classification. Audio acquisition, is the process of collecting information on audio files such as sampling rate and bit depth. The dataset of audio files from traditional Indian string musical instrument were collected from SoundCloud, the online music sharing platform (SoundCloud, 2007). There were a total of 500 audio files since each instrument from the class of Veena contribute to 100 audio files. Audio pre-processing is the process to modify the audio signal based on user preferences. The audio files were cropped when the amplitude value is bigger than or closer to the maximum value which is 0dB (Zytrax, 2014). Meanwhile, audio feature extraction involves the process of extracting low-level features of audio files (Lerch, 2012; Furht, 2009). Three audio feature extraction techniques were used which are Mel-Frequency Cepstral Coefficient (MFCC), Linear Predictive Coding (LPC) and Zero-Crossing Rate (ZCR). Audio classification, on the other hand, is the process to categorize the audio features into a sample of classes in order to obtain their classification accuracy by using different classifiers. Five audio classifiers were used which are k-Nearest Neighbor (kNN), Bayesian Network (BNs), Support Vector Machine (SVM), C4.5 decision tree and Naïve Bayes (NB). In addition, in order to validate the performance of ZFE equalizer, another similar equalizer known as Minimum Mean Square Error (MMSE) equalizer was used in this research (Kumar and Kaur, 2012). MMSE equalizer was integrated with the audio feature extraction techniques in the same way ZFE was integrated with the techniques. The performance of ZFE and MMSE will be compared.

1.5 Importance of Research

The proposed techniques are important because it can be utilized in many areas or field involving soft computing. By studying the techniques used in the audio feature extraction process, the proposed techniques are hoped to provide a better performance of audio classification accuracy when extracting audio signal that contains noise in the highest amplitude. Therefore, this research hopes to solve the problem of audio feature extraction techniques in the domain of Indian traditional instruments.

1.6 Thesis Outlines

For the remaining chapters of the thesis is structured as follows:

Chapter 2 provides a detail explanation of the audio feature extraction techniques used in the research which is Mel-Frequency Cepstral Coefficients (MFCC), Linear Predictive Coding (LPC) and Zero-Crossing Rate (ZCR). Moreover, this chapter also discusses on the sound equalizer which is Zero Forcing Equalizer (ZFE) and Minimum Mean Square Error (MMSE). The benchmark classifiers which are k-Nearest Neighbor (kNN), Bayesian Network (BNs), Support Vector Machine (SVM), C4.5 decision tree and Naïve Bayes (NB) are used to evaluate the performance of audio feature extraction techniques. In addition, the background of traditional Indian musical instrument is mention in this chapter.

Chapter 3 outlines the phases involved in the methodology. The methodology consisted of four phases which is audio acquisition, audio pre-processing, proposed audio feature extraction techniques and audio classification. Besides, this chapter also provides an explanation on the performance measurement in term of classification accuracy, Root Mean Square Error (RMSE) F-Measure and Mean Absolute Error (MAE).

Chapter 4 presents the performance evaluation results based on the classification accuracy, RMSE, F-Measure and MAE of the extracted audio features of audio feature extraction techniques. In this chapter, the comparisons between the original and proposed techniques were mentioned. In addition, a comparison between ZFE and MMSE equalizers is presented. A detail explanation on the confusion matrices is provided.

Chapter 5 concludes and highlights the research finding. This chapter summarizes the research outcome and discusses the advantages of the research. Besides, this chapter provides the future work that can be done to further enhance the research.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter outlines the techniques used in audio feature extraction and audio classification. In addition, a detail description of the techniques were provided. Furthermore, this chapter also provides an explanation on the sound equalizer used in the research. Besides, the overview on the domain of tradition Indian string musical instrument was introduced in this chapter.

2.2 Audio Feature Extraction Concept

Audio feature extraction is a process that involves transforming audio data to a set of features such as pitch, timbre, and others (Bormane & Dusane, 2013; Bullock, 2008). Specifically, audio feature extraction process addresses the analysis and extraction of meaningful information from audio signals. The objective of the audio feature extraction process is to capture the relevant information on audio signal to get a higher-level understanding of the audio signal (Dave, 2013; Bullock, 2008). Furthermore, the extracted features of the audio signal may provide a higher-level understanding of the amplitude or frequency components of the audio signal by plotting to the spectrogram (Dave, 2013; Bullock, 2008). Previously, extracted audio features are obtained by using different type of audio feature extraction techniques such as Mel-Frequency Cepstral Coefficient (MFCC), Linear Predictive Coding (LPC) and Zero-Crossing Rate (ZCR) for musical instrument recognition describing various sound qualities (Bormane & Dusane 2013; Umapathy, Krishnan & Rao, 2007). However, according to Kumari, Kumar and Solanki (2010), the audio features become quite hard to extract robustly when dealing with musical phrases such as bass-line, percussion loops and others. Therefore, choosing the right feature extraction techniques to extract the features is crucial. In addition, the field of music feature extraction is a wide research area, and improving feature extraction will most likely have a major impact on the performance of an instrument classification system (Gunasekaran & Revathy, 2008). In the next section, the research will provide detail descriptions on the chosen audio feature extraction techniques.

2.3 An Overview of Audio Feature Extraction Techniques

There are many techniques that has been used in audio feature extraction such as as Mel-Frequency Cepstral Coefficients (MFCC), Local Discriminant Bases (LDB), Linear Predictive Coding (LPC), Perceptual Linear Prediction (PLP) and Zero-Crossing Rate (ZCR) (Dave, 2013; Anusuya & Katti, 2011; Umapathy, Krishnan & Rao, 2007). In addition, these techniques have their own strengths and drawbacks in extracting the audio signals (Dave, 2013; Keronen *et al*., 2011; Ngo, 2011; Patil *et al*., 2012; Sheetal & Raut, 2012; Umapathy, Krishnan & Rao, 2007). Table 2.1 provides the comparison of audio feature extraction techniques used by researchers. For example, audio feature extraction techniques such as MFCC has proven to be one of the effective techniques used in audio feature extraction which has given a better performance, especially in audio recognition rate (Kumar, Pandya & Jawahar, 2014; Xie, Cao & He, 2012; Keronen *et al*., 2011; Furht, 2009). According to Chachada & Kuo (2014), MFCC features have performed better than LDB features in the domain of artificial (instrumental) and natural sounds. In another study conducted by Dave (2013), the author used MFCC, LPC and PLP audio feature extraction techniques and have stated that compared to MFCC and LPC, PLP is more useful in the domain of speech signals since PLP can discard irrelevant information of speech audio signals whereby improve the speech recognition rate. Furthermore, compared to LDB and PLP, audio feature extraction techniques such as MFCC, LPC and ZCR particularly has been applied many times by the previous researchers in the domain of music instrument, specifically in the traditional Indian musical instrument (Chougule & Chavan, 2014; Bormane & Dusane, 2013; Anusuya & Katti, 2011; Kumari, Kumar & Solanki, 2010; Gunasekaran & Revathy, 2008). Even though, MFCC, LPC and ZCR have shown their good performance, the previous researches have pointed out the drawback of these three establish techniques in extracting noisy signals in the highest amplitude which caused quality degradation in the extracted features and may lead to low classification accuracy (Stulov & Kartofelev, 2014; Dave, 2013; Wolf & Nadeu, 2008; Subramanian, 2006). Detail explanation on techniques is described in the next subsection.

Table 2.1: Comparison of audio feature extraction techniques used by authors

<table>
<tr><th>Author/ Year</th><th>Feature extraction techniques</th><th>Classifiers</th><th>Data</th><th colspan="2">Result</th></tr>
<tr><td>Chachada and Kuo
(2014)</td><td>MFCC
Local Discriminant Bases (LDB)</td><td>K-Nearest Neighbor (kNN)
Support vector machines (SVM)</td><td>37 classes of artificial (instrumental) and natural sounds (Example: Bells, Clapping, Thunder and etc.)</td><td colspan="2">MFCC features have performed better than LDB features by using kNN in the domain of artificial (instrumental) and natural sounds</td></tr>
<tr><td>Kumar, Pandya and Jawahar
(2014)</td><td>MFCC</td><td>K-Nearest Neighbor (kNN)
Bayes Network (BNs)
Support Vector Machines (SVM)</td><td>10 Indian ragas (melody) consisting of 170 tunes</td><td colspan="2">In recognition of raagas (melody), kNN and BNs achieved more than 70% of classification accuracy and SVM achieved more than 80% of classification accuracy by using MFCC</td></tr>
<tr><td rowspan="5">Bormane & Dusane
(2013)</td><td rowspan="5">MFCC
LPC
ZCR</td><td rowspan="5">Wavelet Packet Transform (WPT)</td><td rowspan="5">4 classes of musical Instruments
(Example: Sitar, Piano, Guitar, and etc.)</td><td>Family</td><td>Notes Recognition</td></tr>
<tr><td>String Instruments</td><td>More than 60% of recognition rate</td></tr>
<tr><td>Keyboard Instruments</td><td>More than 50% of recognition rate</td></tr>
<tr><td>Woodwind Instruments</td><td>More than 50% of recognition rate</td></tr>
<tr><td>Brass Instruments</td><td>More than 60% of recognition rate</td></tr>
</table>

Table 2.1 (continued)

<table>
<tr><th>Author/ Year</th><th>Feature extraction techniques</th><th>Classifiers</th><th>Data</th><th>Result</th></tr>
<tr><td>Dave
(2013)</td><td>MFCC
LPC
PLP</td><td>Support Vector Machine (SVM)
Artificial Neural network (ANN)</td><td>Music and Speech Signals</td><td><table><tr><td>Discussion</td></tr><tr><td>Compared to MFCC and LPC, PLP is more useful in the domain of speech signals
ANN is more preferable to be used in the field of speech recognition</td></tr></table></td></tr>
<tr><td>Rocha, Panda and Paiva
(2013)</td><td>Standard audio features (SA) (spectral features and MFCC)
Melodic audio features (MA)</td><td>K-Nearest Neighbor (kNN)
Bayes Network (BNs)
Support Vector Machines (SVM)
C4.5 decision tree
Naïve Bayes (NB)</td><td>903 datasets of emotional music</td><td><table><tr><td>Classifier</td><td>Best result</td></tr><tr><td>BNs</td><td>40% - 62%</td></tr><tr><td>NB</td><td>39% - 48%</td></tr><tr><td>kNN</td><td>40% -60%</td></tr><tr><td>C4.5 decision tree</td><td>34% - 60%</td></tr><tr><td>SVM</td><td>45% - 64%</td></tr></table></td></tr>
<tr><td>Patil
(2012)</td><td>MFCC
LPC
ZCR
Short-Time Energy (STE)
Spectral Feature (SF)</td><td>Support Vector Machines (SVM)</td><td>Humming sounds produced by two male speakers</td><td>The fusion system of MFCC+ZCR+STE+SF gives 86.29% of classification accuracy compared to the MFCC which is 77.71%.</td></tr>
<tr><td>Sheetal and Raut
(2012)</td><td>LPC</td><td>Wavelet Packet Transform (WPT)</td><td>Music and Speech Signals</td><td><table><tr><td>Discussion</td></tr><tr><td>LPC can remove the redundancy in the signal and has the highest rate of audio compression</td></tr></table></td></tr>
<tr><td>Xie, Cao and He
(2012)</td><td>MFCC
LPC
ZCR</td><td>Support Vector Machine (SVM)</td><td>40 samples sound of cutting down trees, sawing trees and trees collapse</td><td>Compared to LPC and ZCR, MFCC can give better performance in audio recognition with more than 80% of recognition rate.</td></tr>
<tr><td>Anusuya and Katti
(2011)</td><td>MFCC
LPC
PLP</td><td>Wavelet Packet Transform (WPT)</td><td>500 samples of clean and noisy Kannada audio signals</td><td><table><tr><td></td><td>Clean signal</td><td>Noisy signal</td></tr><tr><td>MFCC</td><td>70-90%</td><td>50-70%</td></tr><tr><td>LPC</td><td>70-80%</td><td>50-60%</td></tr><tr><td>PLP</td><td>40-60%</td><td>60-70%</td></tr></table></td></tr>
</table>

Table 2.1 (continued)

<table>
<tr><th>Author/ Year</th><th>Feature extraction techniques</th><th>Classifiers</th><th>Data</th><th>Result</th></tr>
<tr><td>Keronen et al.

(2011)</td><td>MFCC</td><td>Gaussian Mixture Model (GMM)</td><td>Recorded audio signals contained noise from public place and car environments</td><td>The performance of MFCC degraded due to the mismatch between the training and recognition noise environments</td></tr>
<tr><td>Ngo

(2011)</td><td>LPC

ZCR</td><td>SVM</td><td>Noisy audio signal captured from the microphone and outputted to the loudspeaker in hearing aids</td><td>ZCR has high signal frequency rate in unvoiced sound compared to voiced sound</td></tr>
<tr><td>Kumari, Kumar and Solanki

(2010)</td><td>MFCC

Auto-correlation</td><td>Artificial Neural network (ANN)</td><td>5 different type of North Indian musical instruments</td><td><table>
<tr><th></th><th>MFCC</th><th>Autocor-relation</th></tr>
<tr><td>Flute</td><td>61%</td><td>79.3%</td></tr>
<tr><td>Dholak</td><td>77.0%</td><td>80.7%</td></tr>
<tr><td>Sitar</td><td colspan="2">60%</td></tr>
<tr><td>Bhapang</td><td colspan="2">64.70%</td></tr>
<tr><td>Instrument family</td><td colspan="2">72 %</td></tr>
</table></td></tr>
<tr><td>Gunasekaran and Revathy

(2008)</td><td>MFCC</td><td>K-Nearest Neighbor (kNN)

Multi-layer perceptron (MLP)

Gaussian Mixture Model(GMM)</td><td>10 Indian musical instrument

(Example: Veena, Sitar, Indian Flute and etc.)</td><td><table>
<tr><th colspan="2">Confidence-based Fusion results</th></tr>
<tr><td>KNN + MLP</td><td>93.6%</td></tr>
<tr><td>KNN + GMM</td><td>92.8%</td></tr>
<tr><td>MLP+GMM</td><td>90.9%</td></tr>
<tr><td>KNN+MLP+GWM</td><td>92.1%</td></tr>
</table></td></tr>
<tr><td>Umapathy, Krishnan and Rao

(2007)</td><td>Local Discriminant Bases (LDB)

MFCC</td><td>Local Discriminant Bases (LDB)</td><td>Artificial and natural sounds (example: drums, flute, animals and etc.)

(213 sounds)</td><td><table>
<tr><th colspan="2">LDB & MFCC (level of average accuracy)</th></tr>
<tr><td>First level</td><td>91%</td></tr>
<tr><td>Second level</td><td>99%</td></tr>
<tr><td>Third level</td><td>95%</td></tr>
</table></td></tr>
</table>

2.3.1 Mel Frequency Cepstral Coefficients (MFCCs)

Mel Frequency Cepstral Coefficients (MFCCs) are cepstral coefficients used for representing audio in a way that mimics the physiological properties of the human auditory system (Kumari, Kumar & Solanki, 2010; Sukor, 2012). The mel scale was developed based on the study of human auditory perception. MFCCs are commonly used in speech recognition and are increasingly used in music information recognition and genre classification systems (Kumari, Kumar & Solanki, 2010). MFCC technique has its own advantages and disadvantages in extracting audio signal. According to Xie, Cao and He (2012), MFCC will give better performance in audio recognition rate. In addition, MFCC will take a short time for extracting the features of audio signal (Keronen *et al*., 2011). However, the drawback of MFCC is it will produce low quality of audio features and leads to low classification accuracy if the signal used is noisy (Anusuya & Katti, 2011). Figure 2.1 shows the block diagram of MFCC.

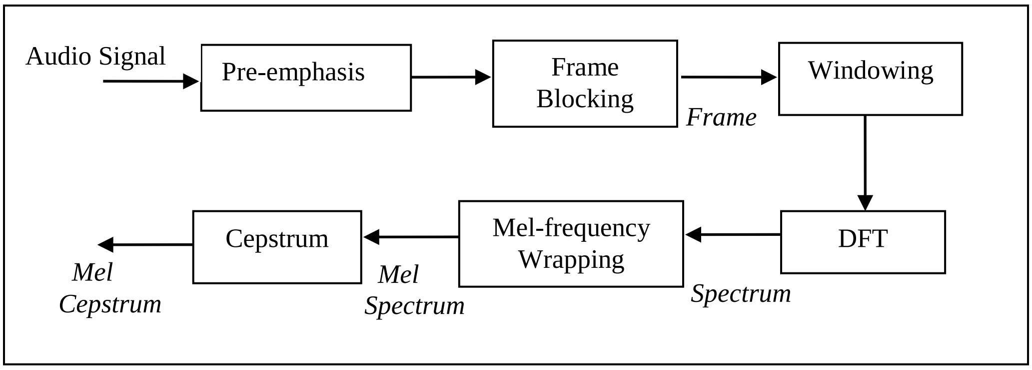

Figure 2.1: MFCC Block diagram (Dave, 2013)

From the MFCC block diagram shown in Figure 2.1, the audio signal from the Indian musical instrument is passed into a process called pre-emphasis. The function of the pre-emphasis process is to boost the energy of signal and amplify the importance of high-frequency formant (Chougule & Chavan, 2014; Le-Qing, 2011; Zhu & Alwan, 2003). Formants are the area in a spectrogram that shows the presence of noise in the highest amplitude by displaying the area as dark bands. Besides, the darker formants produced in the spectrogram shows the audio signals have a stronger energy (amplitude) (Prahallad, 2011; Aslam et al., 2010). The function of the spectrogram is to plot the audio signal in amplitude, frequency or time as shown in Figure 2.2.

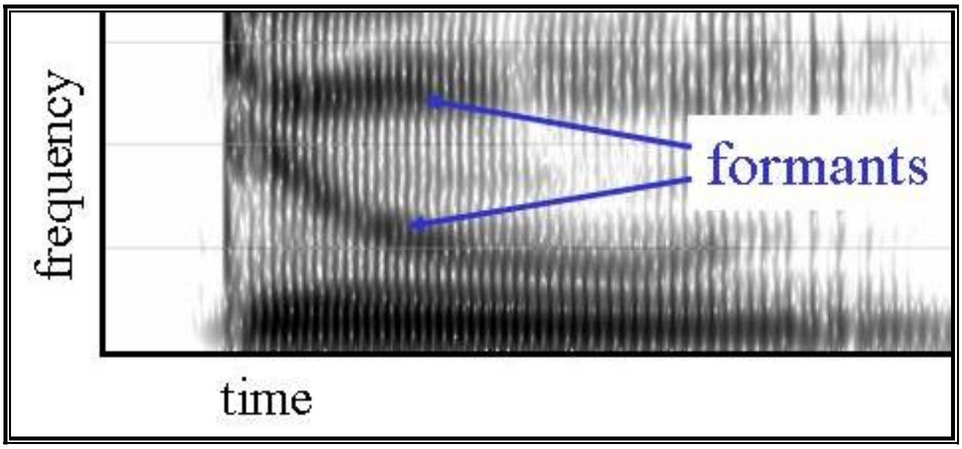

Figure 2.2: Spectrogram diagram (McMurray, 2004)

The equation for pre-emphasis is shown in Equation (2.1):

$$H(z)=1-az^{-1} \tag{2.1}$$

where z is the Discrete Fourier Transform (DFT) of input signal and a is the pre-emphasis alpha coefficient where the value is usually between 0.9 and 1.0 (the standard value is 0.97).

Then, the signal will be passing to the frame blocking where the signal will be blocked into frames of N samples, with adjacent frames being separated by M $(M<N)$. The first frame consists of first N samples; second frame begins with M samples after the first frame, and overlaps it by N-M samples and so on (Gadade, Jadhav & Deogirkar, 2010). The standard parameter or value for M and N is M = 100 and N = 256. The next step is windowing where each individual frame is then windowed by using Hamming window in order to minimize the signal discontinuities at the beginning and end of each frame by taper the signal to zero (Dave, 2013). The Hamming window, $w(n)$ is computed according to the Equation (2.2):

$$w(n) = 0.54 - 0.46 \cos(2\pi n / N\text{-}1),\ 0 \le n \le N - 1 \tag{2.2}$$

where N is total number of sample and n is current sample.

Then, the signal will be processed by using Discrete Fourier Transform (DFT). DFT convert the sampled function from its original domain (often time or position along a line) to the frequency domain. Therefore, each frame of N samples from the time domain is converted into the frequency domain which is defined on the set of N samples $\{x_n\}$, as shown in Equation (2.3):

$$X_k = \sum_{n=0}^{N-1} x_n e^{-j2\pi kn / N}, \qquad k = 0, 1, 2, \ldots, N\text{-}1 \tag{2.3}$$

where X_k's are complex numbers and their absolute values (frequency magnitudes) are calculated. X_k is interpreted as follow: positive frequencies $0 \leq f < F_s / 2$ correspond to values $0 \leq n \leq N / 2\text{-}1$, while negative frequencies $-F_s / 2 < f < 0$ correspond to $N / 2+1 \leq n \leq N\text{-}1$. Here, F_s is considered as sampling frequency.

The output of DFT is defined in spectrum. In the next step, the spectrum is wrapped through a process named mel-frequency wrapping and expressed in the mel frequency scale. According to Dave (2013), mel-frequency scale is linear frequency spacing below 1000 Hz and a logarithmic spacing above 1000 Hz. The output is defined as mel spectrum. The approximate formula to compute the mel is shown in Equation (2.4):

$$m_f = 2595 log_{10} (f / 700 + 1) \tag{2.4}$$

where f is frequency and measured in Hz.

Cepstrum inverses the Fourier transform of the logarithm of the estimated spectrum of an audio signal as shown in Equation (2.5):

$$\hat{C}_n = \sum_{k=1}^{K} (\log \hat{S}_k) \cos[n*(k\text{-}0.5)*\pi/K], \quad n=0,1, ..., K\text{-}1 \tag{2.5}$$

where K is a number of mel spectrum coefficients.

From the equation (2.5), the result is expressed in mel cepstrum and is referred to as the mel-scale cepstral coefficients, or MFCC which is shown in Equation (2.6):

$$C_i(n) = \sum_{m-1}^{M} S(m) \cos[\pi n(m\text{-}0.5)/M] \ , \ \ 0 \leq m < M \tag{2.6}$$

where n is the number of MFCC, $C_i(n)$ is the n-th MFCC coefficients of the i-th frame, $S(m)$ is the logarithmic power spectrum of the audio signal, and M is the number of triangular filters.

2.3.2 Linear Predictive Coding (LPC)

Similar to MFCC, Linear Predictive Coding (LPC) is another technique which offers a powerful, yet simple method to extract audio information. LPC algorithm produces a vector of coefficients that represents a smooth spectral envelope of a temporal input signal (Elminir, ElSoud & El-Maged, 2012). The strength of LPC is it can remove the redundancy in signal. In addition, LPC has the highest rate of audio compression (Sheetal & Raut, 2012) and take short training time, even though it will take long time to extract the feature of audio signal (Wolf & Nadeu, 2008). Nevertheless, using LPC alone for the recognition process is not very successful because the all pole assumption of the vocal cord transfer function was not accurate (Wolf & Nadeu, 2008). On the other hand, LPC coder does not work well for low or high pitch frequency of voices (speech) signal (Kamal, Sarkar & Rahman, 2011). Another importance drawback of LPC is the extracted audio features may lead to low performance of audio classification accuracy if there is noise in the highest amplitude of an audio signal (Chougule & Chavan 2014; Dave, 2013). This is due to LPC synthesizer process amplified the noise in highest amplitude that may decrease the quality of extracted features. Figure 2.3 shows the LPC block diagram.

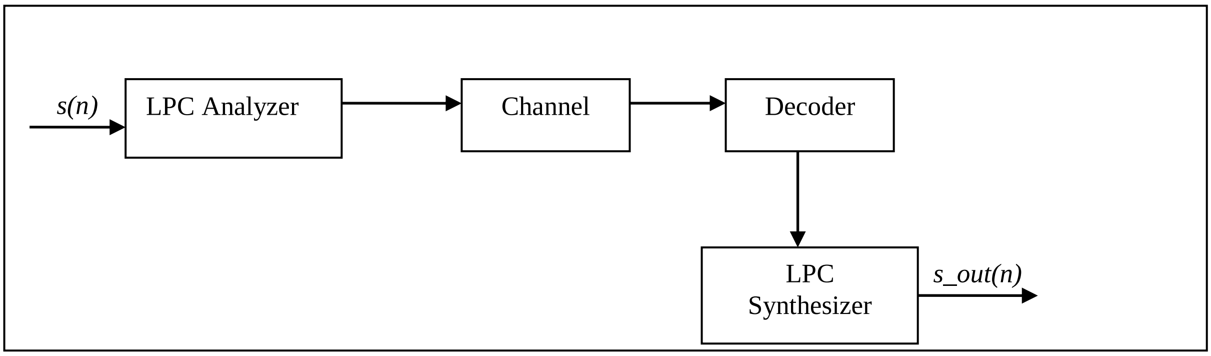

Figure 2.3: LPC Block diagram (Alwan, 2002)

First, the audio signal will be sent to LPC analyzer to determine the key features of the signal and try to encode the signal as accurately as possible. The key features resemble the loudness of the signal determine whether the sound is voiced or unvoiced. Then, the signal is sent to the channel. The function of the channel is to transmit or serve as the medium for transmission. Then, the signal will be passed to a decoder. Decoding involves using the parameters acquired in the encoding and analysis to build a synthesized version of the original audio signal. From the decoder, the signal is passed to LPC synthesizer. The function of LPC synthesizer is as a computerized console or module for creating or scaling the audio signal (Alwan, 2002). LPC synthesizer will scale the output signal to match the level of the input signal (McCree, 2008). In other word, if the input signal contains high energy (amplitude), LPC synthesizer will amplify the output signal to match with the input signal. The formula for LPC synthesizer is shown in Equation (2.7):

$$s(n) = \sum_{k=1}^{p} \alpha_k \, s(n-k) + G\, u(n) \tag{2.7}$$

where $s(n)$ is the waveform samples, α_k is the predictor coefficient, and G represents the loudness and it is multiplied with the excitation signal, $u(n)$ to obtain proper loudness intensity in the excitation signal.

From equation (2.7), the result can be simplified and referred as the Linear Predictive Coding (LPC) which is shown in Equation (2.8) (Aviv & Grichman, 2011):

$$H(z) = G \,/\, (1 + \sum_{k=1}^{p} a_p\,(k)\, z^{-k}) \tag{2.8}$$

where p is the order (number of poles), gain G is the signal loudness, and { $a_p(k)$ } are the input parameters for the audio files.

2.3.3 Zero-Crossing Rate (ZCR)

Zero-crossing rate (ZCR) is the rate at which the signal changes from positive to negative or vice versa. Simply, ZCR is a measure of the number of times in a given frame that the amplitude of the audio signals passes through a value of zero. The rate at which zero crossing occurs is a simple measure of the frequency content of a signal. Figure 2.4 shows the block diagram of ZCR. First, the parameter of audio file such as frame size is assigned. Then, the amplitude of the audio signal is calculated in energy calculation and the output is called ZCR which is shown in Equation (2.9):

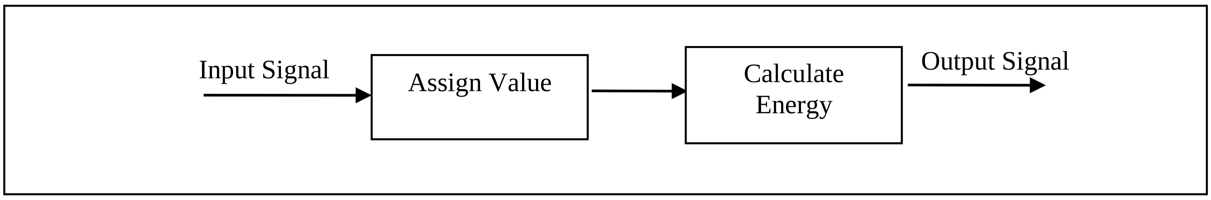

Figure 2.4: ZCR Block diagram (Raju *et al*., 2013)

$$ZCR = (1/T\text{-}1) \sum_{t=1}^{T-1} \|\{\, s_t\, s_{t-1} < 0 \,\} \tag{2.9}$$

where, s is a signal of length T and the indicator function || {A} is equal to 1 when A is true and is equal to 0 otherwise.

In order to use ZCR to distinguish unvoiced sounds from noise and the environment, the waveform can be shifted before computing the ZCR. ZCR has high signal frequency rate and is much lower for voiced sound as compared to unvoiced sound (Ngo, 2011; Patil *et al*., 2012). ZCR is an important parameter for voiced or unvoiced classification (Khan, Bhaiya, & Banchhor, 2012). Unfortunately, the drawback of ZCR is it unable to extract the best quality of audio features due to the energy calculation process that amplify the noise in the highest amplitude of an audio signal (Bormane & Dusane, 2013). Therefore, ZCR leads to degradation of audio classification accuracy (Bormane & Dusane, 2013).

Nevertheless, some improvement can be done to MFCC, LPC and ZCR to improve the performance of the techniques in audio feature extraction whereby contribute to high quality of audio features and lead to a better audio classification accuracy. Therefore, this inspired the use of an equalizer to be integrated with audio feature extraction techniques to equalize the noise in the highest amplitude of audio signal. The next section will provide a detail explanation on the noise in audio signal.

2.4 Noise in Audio Signal

Noise can be defined as uncontrolled, loud, unmusical, or unwanted component of an audio signal (Gold, Morgan & Ellis, 2011; Weeks, 2010; Thorne, 2007). In addition, there are many types of noise such as voiced noise, unvoiced noise or background noise (Gold, Morgan & Ellis, 2011). Voiced noise is the sound produced by the human voice such as breathing or coughing, meanwhile unvoiced noise is the sound produced from the music instrument or sound effects such as door knocks, paper shuffling or plucking a string instrument (Gold, Morgan & Ellis, 2011; Gunasekaran & Revathy, 2008). Background noise is any sound other than the sound being used than can be based on the surrounding or can be considered as an external sounds such as ambulance going by outside or people talking (Gold, Morgan & Ellis, 2011). In this research, the noise has been identified come from the music instrument itself due to the vibration of the string that produced noise in the highest amplitude (Dave, 2013; Stulov & Kartofelev, 2014; Subramanian, 2006). Therefore, some type of noise such as background noise could be reduced or removed by using noise filtering. Specifically, some type of noise such as high amplitude noise can only be mitigated by using sound equalizer (Kaur & Kansal, 2013; Gold, Morgan & Ellis, 2011; Weeks, 2010). The next section will describe on the noise filtering and sound equalizer in more detail.

2.5 Noise Filtering

Noise filtering or also known as noise reduction is a process of removing unwanted components (noise) from an audio signal (Tan & Jiang, 2013). Noise filtering is normally applied in audio signal to perform application such as noise reduction, sound crossover and others (Tan & Jiang, 2013; Altera, 2010). Basically, noise filtering involves the process of removing different types of noise such as background noise or unvoiced noise from the audio signal (Gold, Morgan & Ellis, 2011; Weeks, 2010). Moreover, noise filtering is used to filter a noisy signal, such as cleaning up an audio signal recorded in a room full of other conversations (Tan & Jiang, 2013). Nevertheless, noise filtering does not focusing on the specific high amplitude noise while filtering background noise in the audio signal (Tan & Jiang, 2013; Weeks, 2010). Therefore, as an alternative to the drawback of noise filtering, this study suggests the use of sound equalizer. The sound equalizer has the potential of equalizing the noise that come from the high amplitude of an audio signal (Kaur & Kansal, 2013; Kumar & Kaur, 2012; Altera, 2010). The next section will provide a detail explanation on the sound equalizer.

2.6 Sound Equalizer

Sound equalizer is an essential part of any sound system which provides an approximate inverse of the channel frequency response (Kumar & Kaur, 2012). Equalizers are used in recording studios, broadcast studios, and live sound reinforcement to eliminate unwanted sounds such as noise from microphones, instrument pick-ups, loudspeakers, and hall acoustics. Figure 2.5 shows an audio equalizer in the communication system. In general, the audio captured from the audio input source is sent to an equalizer by using a communication channel (physical medium) such as wires, radio, acoustic, magnetic or optical recording media. The function of an equalizer is to equalize the amplitude of an audio signal and transmit to a digital to analog converter (DAC). DAC will convert the digital audio signal to analog sound and output the analog sound to a speaker or headphone.

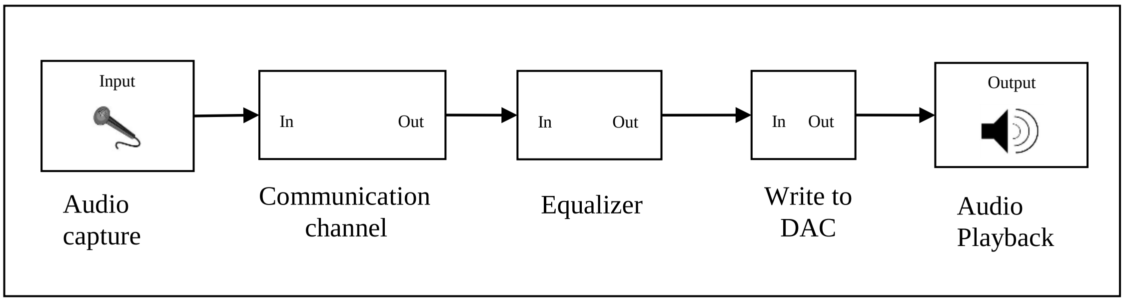

Figure 2.5: Audio equalizer (Proakis, 2008)

Specifically, equalizer involves the process of equalization to mitigate the effects of intersymbol interference (ISI). ISI is a form of distortion of audio signal due to an unwanted sound or effect such as noise which is produced in the high amplitude (Kaur & Kansal, 2013). Kumar and Kaur (2012) also stated that the reduction of ISI effects has to be stabilized since the audio signal contains noise. Based on Deepa (2013), ISI arises because of the spreading of a transmitted pulse due to the dispersive (widely spread or scattered) nature of the channel, which results in overlapping of adjacent pulses. Equalizer is usually implemented at the original frequency or the amplitude range of the signal (National Chung Cheng University, 2013). Equalization equation is shown in (2.10):

$$y(t) = x(t) * f^*(t) + n_b(t) \tag{2.10}$$

where

$x(t)$ is the original input signal

$f^*(t)$ is the complex conjugate of $f(t)$

(Changing the sign of the imaginary part and leaving left real part unchanged)

$n_b(t)$ is the noise at the input of the equalizer

Equalizer responds to the impulse response of the signal. Impulse response refers to the reaction of wave in signal response to some external change (Golden, Dedieu & Jacobsen, 2005). High wave in signal refers to the highest amplitude of wave in the channel. The simplest way to obtain the channel impulse response is to send a single pulse (approximately an impulse) over the channel and observe the resulting received signal. If there is no noise on the channel, this could give a good estimate of the channel impulse response. However, in reality, an estimate of the channel impulse response that is based on a single transmitted impulse will always be noisy. Therefore, the possibility to reduce the effect of the noise is by taking many measurements of the impulse response (Golden, Dedieu & Jacobsen, 2005).

There are two general categories of equalizer which is linear and nonlinear equalizer. Altera (2010) stated that linear equalizer is frequency dependent and it can be highly effective in mitigating the ISI. In summation, a linear equalizer mitigates the ISI of a single audio signal without enhancing the noise. Some examples of linear equalizer are Zero-Forcing Equalizer (ZFE) and Minimum Mean Square Error (MMSE) equalizer. On the other hand, nonlinear equalizer involves the use of mixed-signal (more than one audio signal in the waveform). MIT Lincoln Laboratory (2009) stated that nonlinear equalizer is used to detect small signals in the presence of strong background, such as in radar, signal intelligence, and electronic intelligence systems. According to National Chung Cheng University (2013), some of the example of nonlinear equalizer is Maximum Likelihood Symbol Detection (MLSD) and Maximum Likelihood Sequence Estimator (MLSE). This study will only focusing on linear equalizer that are ZFE and MMSE since linear equalizer highly effective in mitigating the unwanted sound (noise) in the highest amplitude of audio signal (Kaur & Kansal, 2013; Kumar & Kaur, 2012; Altera, 2010). The next subsection will discuss ZFE and MMSE in detail.

2.6.1 Zero Forcing Equalizer (ZFE)

Zero Forcing Equalizer is a linear receiver used in communication systems. This equalizer inverts the frequency response of the channel. Kumar and Kaur (2012) stated that for a channel with frequency response $F(f)$, the zero forcing equalizer $C(f)$ is constructed as shown in Equation (2.11):

$$C(f) = 1 / F(f) \tag{2.11}$$

The output of ZFE is calculated from the numerator and denominator coefficient, which is obtained from the inverse of impulse response in the frequency domain, $F(f)$. Thus, the combination of channel and equalizer gives a flat frequency response and linear phase as shown in Equation (2.12):

$$F(f)C(f) = 1 \tag{2.12}$$

The implementation of ZFE depends on the channel it is used (Proakis, 2008). A channel is used to convey an information signal, for example a digital bit stream, from one or several senders (or transmitters) to one or several receivers. In information theory, a storage device is also a kind of channel, which can be sent to (written) and received from (read). A typical channel is model in discrete domain as shown in Equation (2.13) (Dytso, 2012):

$$y[n] = h[n]*x[n] + z[n] \tag{2.13}$$

where

$y[n]$ is the channel output

$h[n]$ is the channel impulse response

$x[n]$ is the channel input

$z[n]$ is the noise

From equation (2.13), the channel is converted to frequency domain as shown in Equation (2.14):

$$y(f) = h(f)x(f) + z(f) \tag{2.14}$$

ZFE multiplies $y(f)$ and $z(f)$ by $inv(h(f))$ to reduce ISI as shown in Equation (2.15) (Kaur & Kansal, 2013):

$$inv(h(f))y(f) = x(f) + inv(h(f))z(f) \tag{2.15}$$

In previous research, ZFE has been used to solve the problem of signal transaction in Multiple Input Multiple Output (MIMO) systems (Khademi *et al.*, 2013). MIMO technology is a wireless technology that uses multiple transmitter and receiver to improve communication performance. Hence, ZFE is used to mitigate the interference in signal transaction. Moreover, ZFE is much more useful for equalizing the effect of noise in the higher amplitude as introduced by ISI (Kaur & Kansal, 2013; Mobile Communication, 2009). However, the drawback of ZFE is that the channel response may often exhibit attenuation (reduction of signal strength during transmission) at high frequencies around one-half the sampling rate (the folding frequency) (Kumar & Kaur, 2012). Another drawback is that the use of the equalizer as standalone or independently decreases the performance of the channel. Therefore, it depends on how ZFE is used as mentioned by Dytso (2012).

As mentioned previously, due to the drawback of audio feature extraction techniques in extracting noisy signal, ZFE is proposed to be integrated with audio feature extraction techniques since the characteristics of ZFE is to mitigate the effect of noise in the highest amplitude of the audio signal. Hence, ZFE will be able to improve the performance of audio feature extraction techniques. Another equalizer known as MMSE, which is in the same category with ZFE is discussed in the next section.

2.6.2 Minimum Mean Square Error (MMSE) Equalizer

Minimum mean square error (MMSE) equalizer minimizes the mean square error (MSE). MSE is a common measure of estimator quality as stated by Kumar and Kaur (2012). The main function of MMSE equalizer is that it does not usually eliminate ISI completely but it minimizes the total power of the noise in the output. If x is an unknown random variable, then an estimator of x will be any function from the measurement of known random variable, and its MSE is given by the trace of error as shown in the simplified equation (2.16):

$$\mathrm{MSE} = \mathrm{E}\{(\hat{x} - x)^2\} \qquad (2.16)$$

where x is a scalar variable.

According to Cioffi (2008), MMSE can provide better performance if the audio signal is voice (speech). However, the author pointed out the major drawback of MMSE is that the equalizer is slightly more complicated to describe and analyze than the ZFE. Also, because of the biasing (there is an external force that controls the equalizer), the MMSE output is slightly lower than the ZFE output. MMSE does not assume any stochastic mechanism (having random variable) of the desired and observed signals (Chen *et al.*, 2013). It only makes assumptions about the noise. For example, the noise is additive zero-mean, time-independent, bounded (limited), and known variance. It also does not usually reduce the ISI effect. Due to the disadvantages of the MMSE, this research is focusing on ZFE. However, this research will also integrate MMSE with audio feature extraction techniques in order to compare the performance of both equalizers when they are integrated with the respective audio feature extraction techniques.

REFERENCES

Agarwal, P., Karnick, H., & Raj, B. (2013). A Comparative Study Of Indian And Western Music Forms. *In ISMIR*. Brazil: Pontifical Catholic University of Paraná. pp. 29-34.

Altera. (2010). *Understanding the Pre-Emphasis and Linear Equalization Features in Stratix IV GX Devices*. Retrieved on January 29, 2014, from www.altera.com

Alwan, A. (2002). *Wideband Speech Coding with Linear Predictive Coding (LPC).* Retrieved September 13, 2013, from http://www.seas.ucla.edu

Anusuya, M.A. & Katti, S.K. (2011). Comparison of Different Speech Feature Extraction Techniques with and without Wavelet Transform to Kannada Speech Recognition, *International Journal of Computer Applications,* 26(4), pp. 19-24.

Aslam, T., Doherty, J., Dubrau, A., & Hendren, L. (2010). AspectMatlab: an aspect-oriented scientific programming language. *In Proceedings of the 9th International Conference on Aspect-Oriented Software Development*. France: ACM New York. pp. 181-192.

Atlas, L. (2011). *Spectrogram*. Retrieved on March 18, 2015, from http://isdl.ee.washington.edu

Aviv, A., Grichman, K. (2011). *Long-term prediction*. Retrieved on August 30, 2013, from http://health.tau.ac.il

Bello, J. P. (2013). *Sound Classification*. Retrieved on May 2, 2014, from http://www.nyu.edu

Ben Gal, I. (2007). *Bayesian Networks*, in Encyclopedia of Statistics in Quality and Reliability, Ruggeri, F., Kenett, R. S. and Faltin, F. (editors in chief). USA: John Wiley & Sons Ltd.

Bormane, D. S., & Dusane, M. (2013). A Novel Techniques for Classification of Musical Instruments. *In Information and Knowledge Management*, 3(10), pp. 1-8.

Buckingham Music Inc. (2008). *Veenas.* Retrieved on February 4, 2013, from http://www.buckinghammusic.com/veena/veena.html

Bullock, J. (2008). *Implementing audio feature extraction in live electronic music*. University of Birmingham: Ph.D. Thesis.

Chachada, S., & Kuo, C. C. J. (2014). Environmental sound recognition: A survey. *APSIPA Transactions on Signal and Information Processing*, 3(14), pp. 1-15.

Chen, J., Ma, T., Chen, W., & Peng, Z. (2013). Unsupervised robust recursive least-squares algorithm for impulsive noise filtering. *Journal Science China Information Sciences*, 56(4), pp. 1-10.

Cioffi, M.J. (2008). *Chapter 3. Equalization*. Retrieved on January 30, 2014, from http://www.stanford.edu/group/cioffi/book/

Chougule, S. V., & Chavan, M. S. (2014). Channel Robust MFCCs for Continuous Speech Speaker Recognition. *Advances in Signal Processing and Intelligent Recognition Systems*, pp. 557-568.

Chu, S., Narayanan, S., & Kuo, C. C. (2009). Environmental sound recognition with time–frequency audio features. *IEEE Transactions on Audio, Speech, and Language Processing*, 17(6), pp. 1142-1158.

Courtney, D. (2012). *Gottuvadhyam.* Retrieved on February 3, 2013, from http://chandrakantha.com

Dave, N. (2013). Feature Extraction Methods LPC, PLP and MFCC in Speech Recognition. *International Journal of Research in Engineering and Advanced Technology*, 1(6), pp. 1-5.

Deepa,T. (2013). *Linear Equalizers & Nonlinear Equalizers*. Retrieved on January 28, 2014, from http://www.srmuniv.ac.in

Dytso, A. (2012). *Zero Forcing Equalizer*. Retrieved on January 29, 2014, from http://www.mathworks.com

Elminir, H. K., ElSoud, M. A., & El-Maged, L. A. (2012). Evaluation of Different Feature Extraction Techniques for Continuous Speech Recognition. *International Journal of Science and Technology*, 2(10), pp. 689-695.

Eronen, A. (2001). Automatic musical instrument recognition. Tempere University of Technology: Master's Thesis.

Furht, B. (2009). *Handbook of multimedia for digital entertainment and arts*. USA: Springer.

Gadade, M. H., Jadhav, M. M. R., & Deogirkar, M. S. V. (2010). *Speech Identification and Recognition Using Data Mining*. Government College of Engineering: Technical Report.

Gold, B., Morgan, N., & Ellis, D. (2011). *Speech and audio signal processing: processing and perception of speech and music*. USA: John Wiley & Sons.

Golden, P., Dedieu, H. & Jacobsen, K.S. (2005). *Fundamentals of DSL Technology*. USA: Auerbach.

Gunasekaran, S. & Revathy, K. (2008). Fractal dimension analysis of audio signals for Indian musical instrument recognition. *International Conference on Audio, Language and Image Processing (ICALIP).* Shanghai: IEEE. pp. 257-261.

Gupta, S., Jaafar, J., Ahmad, W. F., & Bansal, A. (2013). Feature Extraction Using MFCC. *An International Journal of Signal & Image Processing (SIPIJ)*, 4(4), pp. 101-108.

Jupiter Infomedia Ltd. (2012). *Carnatic music*. Retrieved on December 24, 2012, from http://www.indianetzone.com/2/carnatic_music.htm

Jupiter Infomedia Ltd. (2011).*Veena*. Retrieved on December 24, 2012, from http://www.indianetzone.com/16/veena.htm

Kamal, S. N., Sarkar, R. N., & Rahman, M. (2011). Comparison of Musical Pitch Analysis Between LPC and CELP. *International Journal of Advances in Engineering Sciences*, 1(1), pp. 35-39.

Kaur, N., & Kansal, L. (2013). Performance Comparison of MIMO Systems over AWGN and Rician Channels with Zero Forcing Receivers. *International Journal of Wireless & Mobile Networks*, 5(1), pp. 73-84.

Keronen, S., Pohjalainen, J., Alku, P., & Kurimo, M. (2011). Noise robust feature extraction based on extended weighted linear prediction in LVCSR. *In Proceedings of INTERSPEECH*, Italy: ISCA. pp. 1265-1268.

Khademi, S., Chepuri, S. P., Leus, G., & van der Veen, A. J. (2013). Zero-forcing pre-equalization with transmit antenna selection in MIMO systems. *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver: IEEE. pp. 5046-5050.

Krishnaraj S. (2010). *Carnatic music*. Retrieved on January 7, 2013, from http://www.carnaticindia.com/

Kumar, V., Pandya, H., & Jawahar, C. V. (2014). Identifying ragas in Indian music. *22nd International Conference on Pattern Recognition (ICPR)*. Sweden: IEEE. pp. 767-772

Kumar, M., & Kaur, J. (2012). Performance analysis of BPSK system with ZF &MMSE equalization. *International Journal of Latest Trends in Engineering and Technology (IJLTET)*.1(3), pp. 127-133.

Kumari, M., Kumar, P. & Solanki, S. S. (2010). Classification of North Indian Musical Instruments using. Spectral Features. *Journal of Computer Science and Telecommunications (GESJ)*. 6(29), pp. 11-24.

Le-Qing, Z. H. U. (2011). Insect sound recognition based on MFCC and PNN. *International Conference* on *Multimedia and Signal Processing (CMSP).* China: IEEE. pp. 42-46.

Lerch, A. (2012). *An introduction to audio content analysis: Applications in signal processing and music informatics*. USA: John Wiley & Sons.

Li, T., Ogihara, M., & Li, Q. (2003). A comparative study on content-based music genre classification. *In Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval*. Canada: ACM New York. pp. 282-289.

McCree, A. (2008). *Springer Handbook of Speech Processing*. India: CBS. pp. 97-105.

McMurray, B. (2004). *Experimental Psychology II*. Retrieved on August 20, 2014, from http://www2.psychology.uiowa.edu/

MIT Lincoln Laboratory. (2009). *Nonlinear Equalization for Receiver Dynamic Range Extension*. Massachusetts Institute of Technology: Technical Report.

Mobile Communication (2009). *Equalization, Diversity and Coding Techniques*. Retrieved on August 30, 2013, from http://mc.lctu.cn

Nasridinov, A., & Park, Y. H. (2014). A Study on Music Genre Recognition and Classification Techniques. *International Journal of Multimedia and Ubiquitous Engineering.* 9(4), pp. 31-42.

National Chung Cheng University. (2013). *Equalization, Diversity, and Channel Coding*. Retrieved on January 28, 2014, from web.ee.ccu.edu.tw

Nettleton, D. F., Orriols-Puig, A., & Fornells, A. (2010). A study of the effect of different types of noise on the precision of supervised learning techniques. *International Journal of Artificial intelligence review,* 33(4), pp. 275-306.

Ngo, K. (2011). *Digital signal processing algorithms for noise reduction, dynamic range compression, and feedback cancellation in hearing aids*. University of Leuven: Ph.D. Thesis.

Panda, A. K., & Sahoo, A. K. (2011). *Study of Speaker Recognition Systems*. Georgia Institute of Technology: Ph.D. Thesis.

Peng, L., Yang, D., & Chen, X. (2014). Multi frame size feature extraction for acoustic event detection. *In Asia-Pacific Signal and Information Processing Association, 2014 Annual Summit and Conference (APSIPA)*. Cambodia: IEEE. pp. 1-4.

Patil, H. A., Madhavi, M. C., Jain, R., & Jain, A. K. (2012). Combining Evidence from Temporal and Spectral Features for Person Recognition Using Humming. *Proceedings of the First Indo-Japan conference on Perception and Machine Intelligence*. India: Springer. pp. 321-328.

Prahallad, K. (2011). *Speech Technology*. Retrieved on April 27, 2014, from http://www.docstoc.com/docs/109146955/Topic-Spectrogram_Cepstrum-and-Mel-Frequency-Analysis

Proakis J, Salehi M (2008). *Digital Communications*. 5th edition. New York: McGraw-Hill.

Rahman, R. M., & Afroz, F. (2013). Comparison of Various Classification Techniques Using Different Data Mining Tools for Diabetes Diagnosis. *Journal of Software Engineering and Applications*, *6*(03), pp. 85-97.

Raju, N., Arjun, N., Manoj, S., Kabilan, K., & Shivaprakaash, K. (2013). Obedient Robot with Tamil Mother Tongue. *Journal of Artificial Intelligence*, 6(2), pp. 161-167.

Rao, B. T., Chinnam, S., Kanth, P. L., & Gargi, M. (2012). Automatic Melakarta Raaga Identification Syste: Carnatic Music. *International Journal of Advanced Research in Artificial Intelligence (IJARAI)*, 1(4), pp. 43-48.

Rao, V., Gupta, C., & Rao, P. (2013). Context-aware features for singing voice detection in polyphonic music. In *Adaptive Multimedia Retrieval. Large-Scale Multimedia Retrieval and Evaluation*. Spain: Springer. pp. 43-57.

Rocha, B., Panda, R., & Paiva, R. P. (2013). Music Emotion Recognition: The Importance of Melodic Features. *6th International Workshop on Machine Learning and Music*. Czech Republic: Springer. pp. 1-4.

Sathish, V M. (2010). *South Indian star turns musician for Sharjah show*. Retrieved on February 3, 2013, from http://www.emirates247.com

Sheetal, G. D., & Raut, R. D. (2012). Advance Source Coding Techniques for Audio/Speech Signal: A Survey. *International Journal of Computer Technology & Applications*, 3(4), pp. 1335-1342.

Smith, S. (2013). *Digital Signal Processing: A Practical Guide for Engineers and Scientists: A Practical Guide for Engineers and Scientists*. USA: Newnes.

SoundCloud (2007). Retrieved January 7, 2013, from https://soundcloud.com/

Stulov, A., & Kartofelev, D. (2014).Vibration of strings with nonlinear supports. *Journal of Applied Acoustics*, 76(14), pp. 223-229.

Subramanian, M. (2006). *Carnatic Music and the Computer*. Retrieved on March 12, 2014, from http://www.musicresearch.in

Sukor, A. S. (2012). *Speaker identification using MFCC procedure and noise reduction method*. Universiti Tun Hussein Onn Malaysia: Master's Project Report.

Tan, L., & Jiang, J. (2013). *Digital signal processing: fundamentals and applications*. USA: Academic Press.

The Incredible India Travel (2011). *Musical Instruments*. Retrieved on January 5, 2013, from http://www.theincredibleindiatravel.com

Thorne, R. (2007). *Assessing intrusive noise and low amplitude sound*. Massey University: Ph.D. Thesis.

Umapathy, K., Ghoraani, B., & Krishnan, S. (2010). Audio signal processing using time-frequency approaches: coding, classification, fingerprinting, and watermarking. *EURASIP Journal on Advances in Signal Processing*, 10(1), pp. 1-28.

Umapathy, K., Krishnan, S. & Rao, R. K. (2007). Audio Signal Feature Extraction and Classification Using Local Discriminant Bases. *IEEE Transactions on Audio, Speech & Language Processing*, 15(4), pp. 1236-1246.

Walter, S., Scherer, S., Schels, M., Glodek, M., Hrabal, D., Schmidt, M., & Schwenker, F. (2011). Multimodal emotion classification in naturalistic user behavior. In *Human-Computer Interaction. Towards Mobile and Intelligent Interaction Environments* (pp. 603-611). Springer Berlin Heidelberg.

Weeks, M. (2010). *Digital Signal Processing Using MATLAB and Wavelets*. 2nd ed. USA: Jones and Bartlett.

WEKA (2010). Retrieved February 3, 2013, from http://www.cs.waikato.ac.nz/ml/weka/

Witten, I., Frank, E., & Hall, M. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. 3th edition. USA: Morgan Kaufmann.

Wolf, M. & Nadeu, C. (2008). Evaluation of different feature extraction methods for speech recognition in car environment. *15th International Conference on Systems, Signals and Image Processing (IWSSIP)*, Bratislava: IEEE. pp. 359-362.

Xie, C., Cao, X., & He, L. (2012). Algorithm of Abnormal Audio Recognition Based on Improved MFCC. *International Workshop on Information and Electronics Engineering (IWIEE)*, China: Elsevier. pp. 731-737.

Yu, Y., Zimmermann, R., Wang, Y., & Oria, V. (2013). Scalable Content-Based Music Retrieval Using Chord Progression Histogram and Tree-Structure LSH. *IEEE Transactions on Multimedia*, 15(8), pp. 1969-1981.

Zhu, Q., & Alwan, A. (2003). Non-linear feature extraction for robust speech recognition in stationary and non-stationary noise. *Journal of Computer Speech & Language*, 17(4), pp. 381-402.

Zytrax (2014). Digital Sound Primer. Retrieved on August 27, 2014, from http://www.zytrax.com