HYBRID MFCC AND LPC FOR STUTTERING ASSESSMENT USING NEURAL NETWORK

By

CHOO CHIAN CHOONG

A Dissertation submitted for partial fulfilment of the requirement for the degree of Master of Science (Electronic Systems Design Engineering)

March 2016

Acknowledgement

I would like to take this opportunity to thank and appreciate those who supported me either through technically, mentally or physically, those who made this thesis possible. Special thanks to my supervisor and thesis advisor, Dr. Syed Sahal Nazli Alhady Syed Hassan, for his support and guidance. Without his help, his motivation, and his technical knowledge, this thesis wouldn't be completed. The time he spent on this project are very much appreciated.

Last and not least, I want to thank to my wife, Sim Ai Jin, my family and all my fellow friends on the moral supports and help throughout this study.

Table of Contents

List of Tables

List of Figures and Illustrations

List of Abbreviations and Nomenclature

Abbreviation	Meaning
ANN	Artificial Neural Network
AR	Autoregressive
DCT	Discrete Cosine Transform
FFT	Fast Fourier Transform
HTK	Hidden Markov Model Toolkit
k-NN	K-Nearest Neighbor
LD	Lexical Disfluencies
LDA	Linear Discriminant Analysis Classifier
LPC	Linear Prediction Coefficient
LPCC	Linear Prediction Cepstral Coefficient
LS SVM	Least Square Support Vector Machine
MFCC	Mel Frequency Cepstral Coefficient
P	Prolongations
PLP	Perceptual Linear Prediction
R	Repetitions
SD	Supralexical Disfluencies
SFS	Speech Filing System
SLP	Speech Language Pathologist
SVM	Support Vector Machine
UCLASS	University College London's Archive Of Stuttered Speech

Abstrak

Kegagapan merupakan sejenis gangguan ucapan, di mana ia menghalang seseorang cakap dengan fasih. Penilaian terhadap kegagapan dengan cara tradisional memakan masa dan hasilnya mungkin berbeza bagi pakar penilaian yang berbeza. Sistem penilaian kegagapan akan mengurangkan masa dan hasil penilaian yang lebih consisten. Objektif projek ini adalah untuk membina sistem klasifikasi untuk pemanjangan dan pengulangan dalam ucapan yang gagap dengan menggunakan rangkaian neural. Tiga ciri pengekstrak telah digunakan dalam projek ini, iaitu MFCC, LPC dan hibrid MFCC dan LPC. Aliran projek ini adalah: 1) memperoleh data ucapan gagap; 2) segmentasi perkataan dan kategori; 3) pengekstrak ciri dengan menggunakan 3 cara yang berbeza; 4) Klasifikasi menggunakan corak pengiktirafan neural dalam Matlab. Ketepatan keseluruhan menggunakan 3 beza pengekstrak ciri dalam ANN adalah 84.6% (LPC), 84.6% (MFCC) dan 88.5% (MFCC hibrid dan LPC). Ketepatan klasifikasi terhadap kelas yang beza, pemanjangan, pengulangan dan fasih adalah 66.7%, 92.3% dan 96.3% (hibrid MFCC dan LPC). Sistem klasifikasi pegagapan telah dibina, dengan hibrid MFCC dan LPC sebagai pengekstrak ciri, dan ANN sebagai agen klasifikasi. Keseluruhan ketepatan kalsifikasi adalah 88.5%.

Abstract

Stuttering is characterized by disfluencies, which disrupt the flow of speech. Traditional way of stuttering assessment is time consuming. The stuttering assessment results always inconsistent between different judges, because human perception on the stuttering event are different for each individual. The stuttering assessment system will reduce the tedious manual work and improve the consistency of the assessment result. The objective of this project is to develop classifier for prolongation and repetition disfluencies in speech using artificial neural network. Three different feature extraction was used in this project, which is Mel Frequency Cepstral Coefficient (MFCC), Linear Prediction Coefficient (LPC) and hybrid MFCC and LPC. The flow of the project were: 1) Stuttered speech data acquisition; 2) Word segmentation and categorization; 3) Feature extraction using 3 different methods; 4) Classification using neural pattern recognition in Matlab. The overall accuracy of the 3 different feature extraction used were 84.6% (LPC), 84.6% (MFCC) and 88.5% (hybrid MFCC and LPC). The classification accuracy using hybrid MFCC and LPC with respect to target classes, which were prolongation, repetition and fluent, were 66.7%, 92.3% and 96.3%. A disfluencies classifier had been developed with hybrid MFCC and LPC as feature extraction and ANN as a classifier. The overall performance of the disfluencies classifier is 88.5%.

CHAPTER 1

INTRODUCTION

1.1 Overview

There is about 1% people have stuttering in a social group, and about 6% to 6.6% of children have stuttering. 80% of stutterer will be automatically cured. While 1% of those remaining 20%, the chances of curing from stuttering is very low, and they become “developmental stuttering”.

Stuttering is characterized as disfluency. Stuttering is a speech disorder, where the speech performance is not fluent. The feature of stuttering including pronunciation repetition, prolonged, blocked or stalled at the syllable or phone level (Ooi and Yunus, 2006, Tian-Swee et al., 2007, Czyzewski et al., 2003). Prolongations are the events that extend speech segments. Pause is the most well-known disfluencies, which interrupt the rhythmic flow of language (World Health Organization, 2006).

Stuttering happens when the forward flow of speech is interrupted abnormally by prolongations or repetitions of a sound, syllable, or articulatory posture, or by avoidance and struggle behavior. Stuttering, is a disorder in speech rhythm, in which the individual knows precisely what he wishes to say but unable to say due to an involuntary repetition, prolongation, or cessation of a sound (Hedge, 1998).

Stuttering can be categorized as follows (Zang, 1991):

a) Bursts stuttering: Repeated of a syllable when speaking.
b) Reciprocating stuttering: Repeated some syllables when speaking.
c) Blocking stuttering: Sentence at the beginning of words may still be barely able to pronounce. However, when meeting the hard pronunciations, their words are blocked.
d) Ankylosing stuttering: If a few bursts stuttering occurred, they will be nervous, tongue and just like froze. Even the easy words difficult be pronounced.
e) Difficult pronunciation stuttering: The first pronunciation of every sentence will not be able to say. Despite they try hard, they can only pronounce in a dull, low "ah", "eh" sound.

There are 2 broad classes of speech disfluencies, which can be grouped as lexical disfluencies, LD (disfluencies involved single words) and supralexical disfluencies, SD (disfluencies that occur over groups of words).

1.2 Problem Statements

The stuttering assessment is done by counting and classifying the occurrences of disfluencies such as repetition and prolongation by the speech language pathologist (SLP), which is the conventional method to be used. This method is somehow subjective, inconsistent, time consuming and prone to error (Howell et al., 1997, Nöth et al., 2000, Ravikumar et al. 2008). Manual counting of speech disfluencies is

subjective, due to the nature of the individual human judge, which might be different for each speech therapist. Hence, an automatic stuttering system is needed to improve the assessment for more accurate of quantitative and qualitative assessment, in terms of consistency and reliability without any human error in assessment. This automatic stuttering assessment research can be useful when artificial intelligence system is created for speech recognition. This can be also used in the automatic diagnosis of the type of disfluency or therapy (Ooi and Yunus, 2006).

1.3 Project Objectives

The aim of the research is to develop a classifier for prolongation and repetition disfluencies in speech. In order to achieve this aim, the following objectives are adopted:

i. To investigate and evaluate feature extraction methods, for the extraction of speech analysis features.
ii. To investigate and evaluate hybrid feature extraction method to produce input features for a stuttering classifier.

1.4 Scope of Project

The scope of this research work includes the development of a classifier for prolongation and repetition disfluencies in speech. The scope of this project mainly

classifies for LD element, specifically Ps and Rs, in a recorded stuttered speech. This project will only target the stuttering assessment based on the English language.

1.5 Project Outline

The Outline of the report is organized as follows:
Chapter 1 is the introduction chapter, which including the background, and objective and scope of the study.

Chapter 2 is on literature review describes the generic characteristics of stuttering assessment system. Previous workings related to the study are being analyzed and evaluated in this chapter as well. Techniques of classification and feature extraction used to classify speech disfluencies are of particular interest for this research.

Chapter 3 will cover the methodology used to build the classifier for speech stuttering assessment automation in this research. This will be including the methodology used and explains the choice of selection for the particular method.

Chapter 4 will cover the result and discussion of this research. This chapter will describe the performance of the classifiers investigated in this study on the task of categorizing speech disfluencies. Discussion will be made on the result, and further improvement to produce a better classification of speech disfluencies will be mentioned in this chapter as well.

Chapter 5 will conclude the research. Limitation of the methodology used in this research will be discussed. Future works and improvement in this field of study will be discussed in the chapter.

CHAPTER 2

LITERATURE REVIEW

This chapter includes the background study regarding speech stuttering assessment, previous projects, journals and thesis. It also discusses on the components that are used in the project.

2.1 Stuttering Assessment

Stuttering is characterized as disfluencies, which is a disruption of the flow of speech. Among the best known disfluencies include repetitions, prolongations and pauses (Czyzewski et al., 2003). Repetitions refer to the segments of speech, including sounds, syllables and words that are repeated. Prolongations are event in speech, which the speech segment is extended. Pauses is an interruption to the rhythmic flow of language (World Health Organization, 2006).

There are few symptoms of stuttering, which include the fluency of the speech is intermittent impaired. The stutterer is difficult to find the correct words or phrases to express himself. Intermittent impairment of speech fluency, amplitude and loudness can be observed in the event of stuttering. When stuttering happened to a stutterer, their facial expression and posture show the difficulty of continuing to sentences.

Speech motor disorders or voice problems are also the main reason of stutter. The youngest stutterer is about 18 months, when speech emerges. The incidence of stuttering (which is the approximate percentage of the population who have stuttered at any time in their lives) can be as high as about 10% (Peter, 2002).

Stuttering can happened without control. Other movements, or negative emotion, such as embarrassment, fear or irritation can be observed during the event of stuttering (Mulligan et al., 2001).

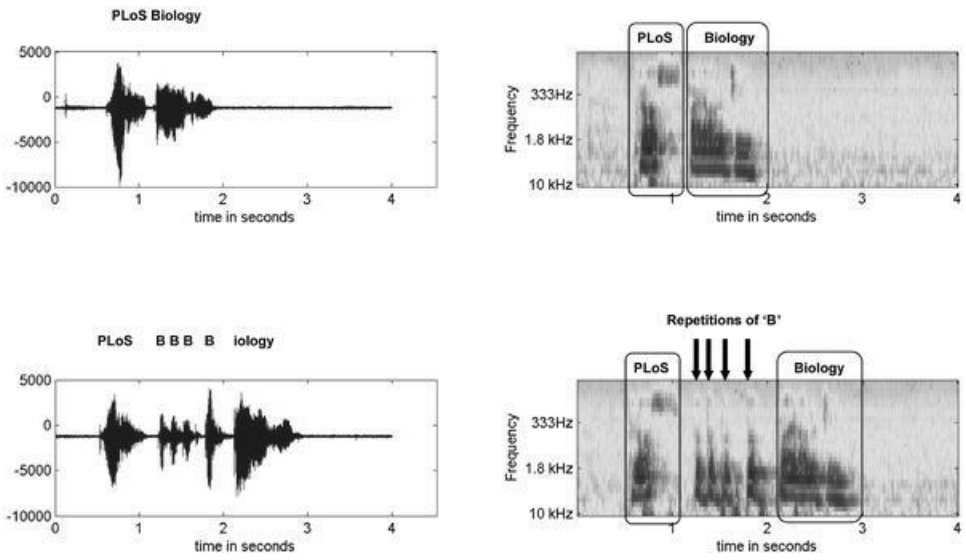

Figure 2.1 Speech Waveforms and Sound Spectrograms of a Male Client Saying "PLoS Biology" (Ooi, 2007)

Figure 2.1 shown the speech waveform sample of a normal speech and stuttered speech. Top row shown the fluent speech and the bottom row shown stuttering typical repetitions occur at the "B" in "Biology". Repetition is happening and clearly identified (shown with arrows) in the spectrogram.

2.1.1 Formal Measures of Assessment System

Speech is a dominant factor in stuttering assessment system. Scales, measuring on a variety of factors, and questionnaires on interpersonal communication are often used by the researchers and SLP when they are doing the stuttering assessment of their client.

According to Erickson, person who stutter have different attitudes compare to those non-stutterers. These formal measures of the interpersonal communication questionnaires and stuttering assessment are able to assess whether a person is having stuttering or not (Walter, 2001).

Several assessment instruments may use to obtain a formal measurement of stuttering, which are:

1. Stuttering Severity Instrument (SSI-3) (Julie et al., 2005)
2. Modified Erickson Scale of Communication Attitudes (S-24) (Yaruss et al., 2006)
3. Perception of Stuttering Inventory (PSI) (Walter, 2001)
4. Locus of Control Behavior (LCB) (Yaruss et al., 2006)
5. Crowe's Protocols (Crowe, 2000)
6. Communication Attitude Test-Revised (CAT-R) (DeNil and Brutten, 1991)
7. A-19 Scale for Children Who Stutter (Kaitlyn, 2004)

2.2 Problems of Classical Manual Stuttering Assessment System

Many researches are discussing on the reliability and validity lack of traditional assessment. Research (Davis et al., 2000) has repeatedly mentioned that certain manual stuttering assessment tools are unreliable. This is because manual assessment totally depends on the SLP's perception on each client's performance.

It is commonly found that in manual measuring techniques that are traditionally used in the clinics, different SLP is having different results when the estimate stuttering event on the identical speech samples (Nail, 2005).

Another problem with classical manual stuttering assessment tools are time consuming. Shorter time spent by the stutterer and SLP in clinic is preferable. The problems could be solved if a computer-based classification system are available.

2.3 Discussion for stuttering assessment system

There are many research works done in the few decades to help SLP during stuttering assessment. Most of the researchers are addressing stuttering recognition system according acoustic analysis, parametric and non-parametric feature extraction, automatic pattern recognition or statistical methods.

In this section, few stuttering assessment system projects will be discussed and compared. The general stuttering assessment working flow can be categorize into a

few parts, which are speech preprocessing, segmentation, feature extraction and classification. These will be discussed in the next section.

2.3.1 Stuttering Speech Data Source

First, the stuttering speech sample data are needed to get things started. Most of the stuttering assessment researches are getting the speech sample from University College London's Archive of Stuttered Speech (UCLASS) database, in http://www.uclass.psychol.ucl.ac.uk/ site (Howell et al., 1997, Lim et al., 2009, Hariharan et al., 2012, Ravikumar et al., 2009, Palfy and Pospichal, 2011). UCLSS database is free access and aiming for research and teaching purpose. Some researchers do have their own stuttering disfluent patients to record the stuttered speech (Ravikumar et al., 2009, Palfy and Pospichal, 2011). Speech sample size is recommended to be more than 100, which are found most of the researchers are using more than 100 with different individuals and genders (Lim et al., 2009, Hariharan, 2012). This is to make sure the system is more robust and stable. In this research, data from the UCLASS database will be used since this is the most reliable speech stuttering data, and commonly used by the researchers in this field.

2.3.2 Speech Segmentation

Stuttered speech needs to segment into small pieces, break into speeches with the small duration of time. There are several of segmented speech duration used in the previous projects, example, Palfy, J et al. were done the segmentation with a time frame of 0.8 s. This segmented speeches will then categorize into a few categories, fluent or dysfluent. The most direct way of doing segmentation is segment the speech samples manually by SLP, this methodology is found in a few projects (Lim et al., 2009, Ravikumar et al., 2009). Stuttering events (Repetitions and prolongations) were identified by the SLP and segmented manually after analyzing the recorded speech. This segmentation method is straightforward, but tedious. Automatic syllable segmentation techniques are available, such as first Autoregressive (AR) coefficient, etc., which mention in the papers (Ravikumar et al., 2009).

Segmented stuttered speech will then be categorized for the purpose of the training for the classifier. There are normally categorized into few category pair:

a) Fluent and dysfluent

b) Fluent, prolongations and repetition

c) Repetitions and prolongations

The database provided by the UCLASS had manually segmented and categorized with 3 categories, fluent, prolongations and repetition. This research used the UCLASS database with the manual segmented and categorized stuttered speech data, for the ease of the research purposes.

2.3.3 Speech Data Preprocessing

The segmented and categorized stuttered speech data will then go through the speech data preprocessing. Preprocessing of the speech is normally done before any speech recognition or speech processing or analysis being done. Digitized speech waveform normally has a wide dynamic range and additive noise. Noise mention here refers to the ambient noise when the recording is being made. Lim et al had used the same technique of reemphasize in both researches. The speech samples were down sampled to 16 kHz. A z-transform filter is pre-emphasized with the following filter.

$$H(z) = 1 - a * z^{-1} \;\; 0.9 \leq a \leq 1.0 \tag{2.1}$$

The purpose of pre-emphasis is to spectrally flatten the signal, so that the speech processing is lesser susceptible to finite precision effects (Venkateswarlu et al., 2012). In the time domain, the filter can be shown as follows. For both researches, Lim et al were using 0.97 for "a" value .

$$s'_n = s_n - as_{n-1} \tag{2.2}$$

Hariharan et al. down sampled the speech form the UCLASS (44.1 kHz) to 8 kHz, since the most of the salient features for speech processing are within 8 kHz bandwidth.

2.3.4 Feature Extraction

In theory, speech recognition can be done directly using the digitized waveform of the speech. The large variable of the digitized speech signal degrade the efficiency of speech recognition process. Feature extraction is to reduce those variability or parameters of the digitized speech signal, for faster and more accurate speech

recognition. Feature extraction is the technique to compute compressed feature vectors that represent the input speech signal (Ghile, 2015). Feature extraction involve of the process for eliminating unnecessary or redundant information, while retaining only the important information for the speech recognition recognizer. The outcome of the feature extraction is to generate a set of necessary parameters, which are just enough to represent a speech signal. Most features extraction package produce a multidimensional feature vector for every speech segment.

Before those automatic feature extraction techniques were invented, the feature extraction was done in a manual way. In Howell et al. Automatic stuttering assessment works, feature extraction of the signals were done manually with 9 parameters. 9 parameters are extracted from the speech, which are: whole word and part word duration; whole word, first part and second part fragmentation; whole word, first part and second part spectral measure; and part word energy.

There are tons of the feature extraction techniques for speech, which are listed in the paper, Techniques for feature extraction in speech recognition system: a comparative study (Shrawankar, 2013). Several feature extraction methods were commonly found and used in the stuttering assessment system, which are Linear Prediction Coefficient (LPC), Mel Frequency Cepstral Coefficient (MFCC) and Perceptual Linear Prediction (PLP) Cepstral.

2.3.4.1 Linear Prediction Coding/Coefficient (LPC)

Linear prediction coding is also called as linear prediction coefficient. This is one of the tools that used mostly in the speech signal analysis. LPC was first proposed as a method for encoding human speech by the United States Department of Defence in federal standard 1015, published in 1984. LPC is one of the most powerful speech analysis techniques, it's a convincing method for encoding quality speech at low bit rate (Shrawankar, 2013). This feature extraction technique able to get the compressed parameters of speech, which represent the spectral envelope of a digital speech signal. LPC uses the information of a linear predictive model to get the parameters. LPC is one of the compression methods that models the speech production process. LPC models this process as a linear sum of the earlier samples using a digital filter inputting an excitement signal (Jeremy, 2000). The basic idea of LPC is a specific speech sample at the current time can be approximated as a linear combination of past speech samples (Ghile, 2015). In other words, linear prediction filters attempt to predict future values of the input signal based on past signals. The disadvantages of the LPC are a small error will make the filter unstable and distort the whole spectrum (Venkateswarlu et al., 2012).

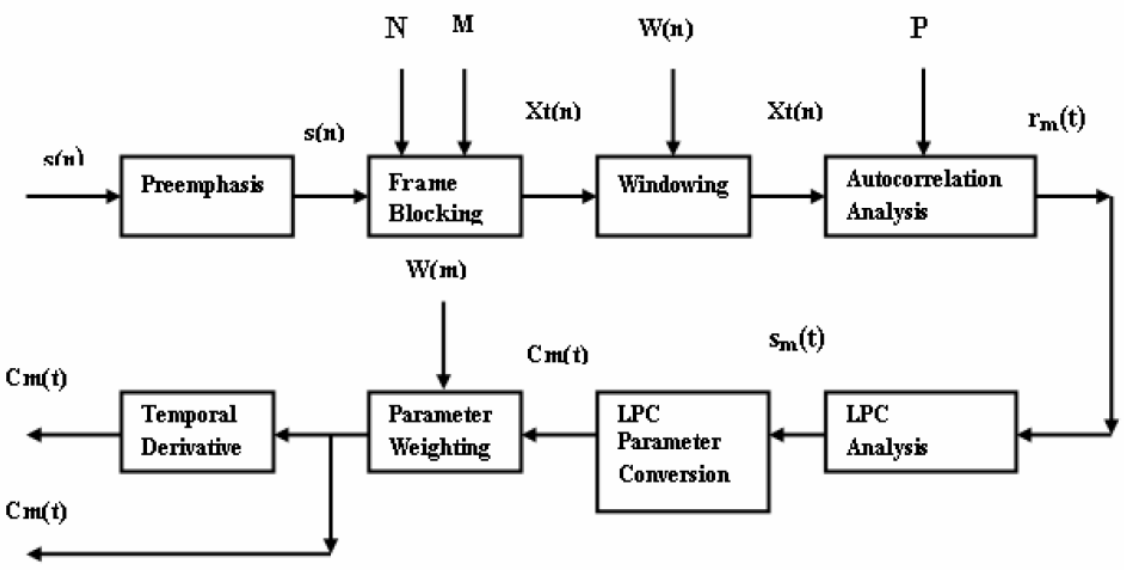

Figure 2.2 The LPC Feature Extraction Technique

Figure 2.2 shown the LPC analysis process (Venkateswarlu et al., 2012). First, the speech signal will go through pre-emphasis process. The pre-emphasis process had discussed in the previous section 2.3.4.

Frame blocking is the process when a speech signal is to be blocked into N samples of frames, with the adjacent frames begin separated by M samples. If $M \leq N$, then LPC spectral estimates from the frame to frame will be smooth, since there are overlaps between adjacent frames, and vice versa.

Windowing will be done on each frame to reduce signal continuities. The signal is tapered to zero at the starting and ending of each frame. The windowed signal can be expressed as below, where w(n) is the window:

$$\tilde{x}(n) = x(n)w(n), 0 \leq n \leq N-1 \quad (2.3)$$

Hamming window is used commonly, which shown as below:

$$w(n) = 0.54 - 0.46\cos\frac{2n}{N-1}, 0 \leq n \leq N-1 \quad (2.4)$$

Autocorrelation analysis is used to find the fundamental frequency. This technique is used to identify both missing fundamental frequency and repeating patterns of the signal. It relies on searching relationship between the signal and a delayed signal of itself. The next auto correlated for each frame of the windowed signal are given in the equation below:

$$R(n) = \sum_{n=0}^{N-1-m} \tilde{x}(n)\tilde{x}(n+m), m = 0,1,2,..,P. \quad (2.5)$$

Where the highest autocorrelation value, P is the order of the LPC analysis. Selection of P depends on the sampling rate.

LPC analysis will be done in the next step. The process will convert each frame of autocorrelation coefficients R into LPC parameters or can be called as LPC coefficients as well. The method is called as Durbin's method. Levinson-Durbin recursion will be utilized to compute the LPC parameters.

$$E_0 = R(0)$$

$$ki = [R(i) - \sum_{j-1}^{i-1} a_j^{i-1} R(i-j)] \,/\, E_{i-1}, 1 \leq i \leq p$$

$$a_i^i = a_i^{i-1} - k_i a_{i-j}^{i-1}, 1 \leq j \leq i-1$$

$$Ei = (1 - k_1^2)E_{i-1} \quad (2.6)$$

The above set of equations (2.6) is solved recursively for i = 1,2,…p, where p is the order of the LPC analysis. The k_i are the reflection or PARCOR coefficients. LPC coefficients, a_j is given as

$$a_j = a_j^{(p)}, 1 \leq j \leq P \quad (2.7)$$

Linear Prediction Cepstrum Coefficients (LPCC) is the extension to LPC. This is the Linear Prediction Coefficients (LPC) represented in the cepstrum domain. LPCC has been proven that it is more reliable and robust than LPC (Lim Sin Chee, 2009).

2.3.4.2 Mel Frequency Cepstral Coefficient (MFCC)

MFCC was introduced by Davis and Mermelstein in the 1980's. Most automatic stuttering assessment systems employ MFCC method for feature extraction. MFCC is a commonly known feature extraction method for speech recognition (Lim Sin Chee et al., 2009, Ravikumar et al., 2009, Palfy et al., 2011). Memon et al showed that human perception on speech recognition does not follow a linear scale, but follow the frequency contents of sounds of speech signal, according to a psychophysical study (Memon et al., 2009). Figure 2.3 shows the block diagram of the MFCC extraction algorithm (Lim et al., 2009).

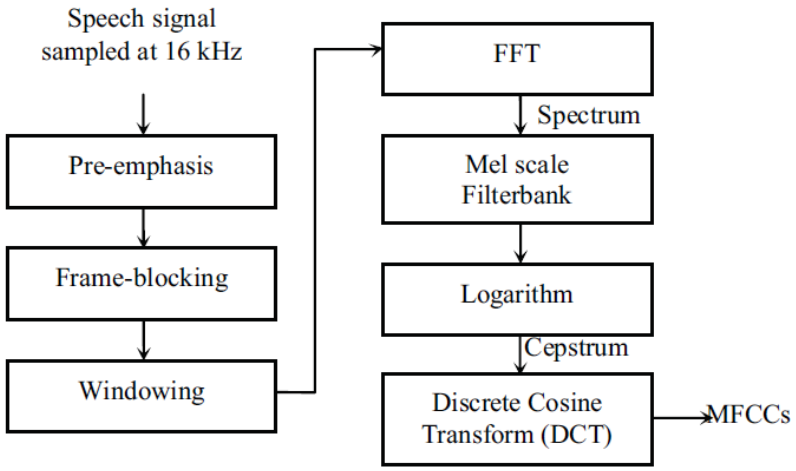

Figure 2.3 Block diagram of MFCC

The high level of the MFCC implementation is (Lim et al., 2009):

1. The speech signal is framed and windowed into short frames/blocks
2. Periodogram estimation of the power spectrum is calculated for each block
3. Mel filterbank is applied to the power spectra, energy for each filter is summed
4. Logarithm for all the filterbank energies
5. Discrete cosine transform (DCT) for the log filterbank energies
6. DTC coefficient 2-13 will be taken as a MFCC output vector.

Compared with LPC, the first 3 steps of MFCC implementation are identical, which are pre-emphasized, frame-blocking and windowing. The framing of the signal is normally done around 20-40 ms per frame. Shorter frame will cause the spectral estimation not reliable, due to insufficiency of the samples. Longer frame will cause too much of changes throughout the frame.

The power spectrum will be calculated on each frame. The power spectrum is in time domain. Fast Fourier transform will be done on the power spectrum in order to get the power spectrum in the frequency domain. The power spectrum in the frequency domain is often called as periodogram. Example of FFT is shown as below:

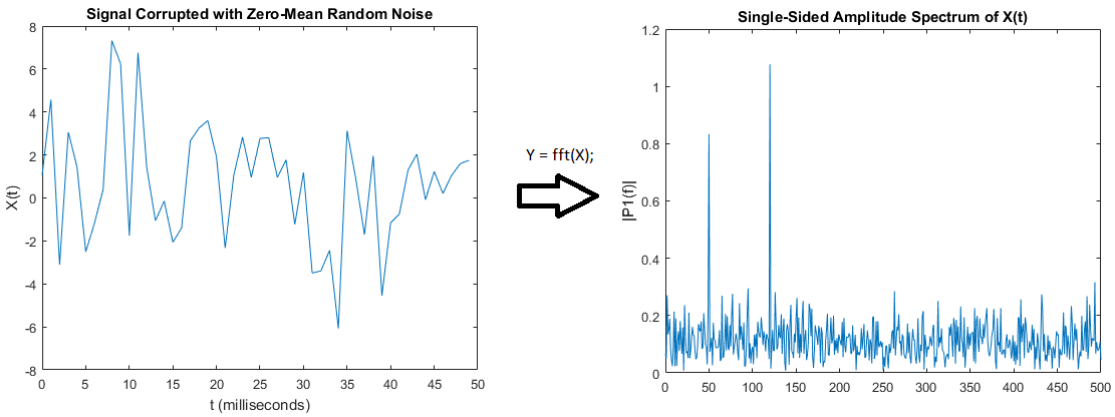

Figure 2.4 Fast Fourier Transform Example

FFT converts the power spectrum from time domain into the frequency domain. Lim et al. used 512 point FFT to obtain a high resolution signal spectrum. Then the power spectrum will be filtered with Mel filterbank, which also called Mel scale spaced filterbank. Mel filterbank is a set of triangular filter spaced on Mel-scale. This mimic the human auditory system's response, where humans are more sensitive to low frequency comparing to high frequency (Ravikumar et al., 2009). The example of the mel scale spaced filterbank is shown as below:

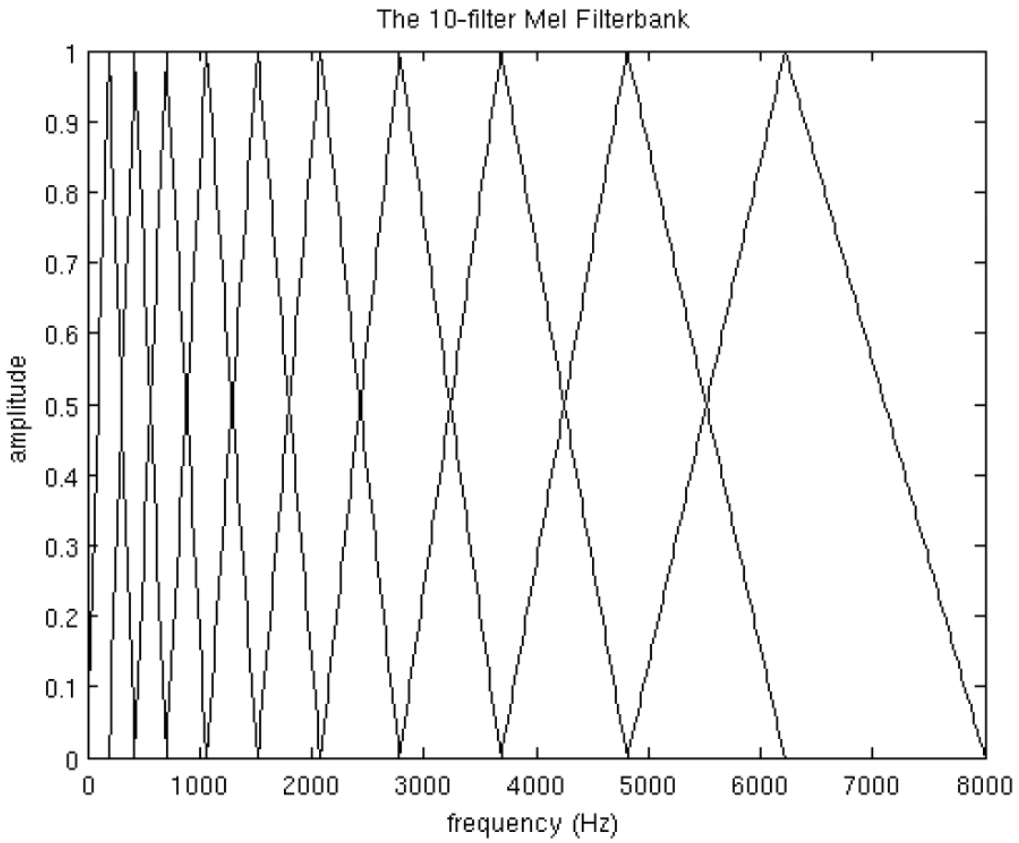

Figure 2.5 10-Mel Space Filterbank in the range of 0 Hz to 8000 Hz

Mel-scale is defined as a logarithmic scale of frequency based on human pitch perception (Lim, 2009). Mel-scale equivalent value for frequency f expressed in Hz is shown as below (O'Shaughnessy, 1987):

$$Mel(f) = 2595\, log_{10}(1 + \frac{f}{700}) \tag{2.8}$$

The example showed a 10-mel space filterbank. For speech recognition purposes, 26-mel space filterbank were commonly used (Lim Sin Chee et al., 2009, Ravikumar et al., 2009, Palfy et al., 2011).

The power spectrum multiply with 1 triangular filter in the Mel filterbank, and then sum up the amplitude with the coefficient. These will resulting the energy in 1 filter. The same step is done for the next 26 filter in the Mel filterbank. In a 26 Mel

filterbank, 26 results will be obtained, which represents how much energy in each filter. Detail of calculation is shown as below:

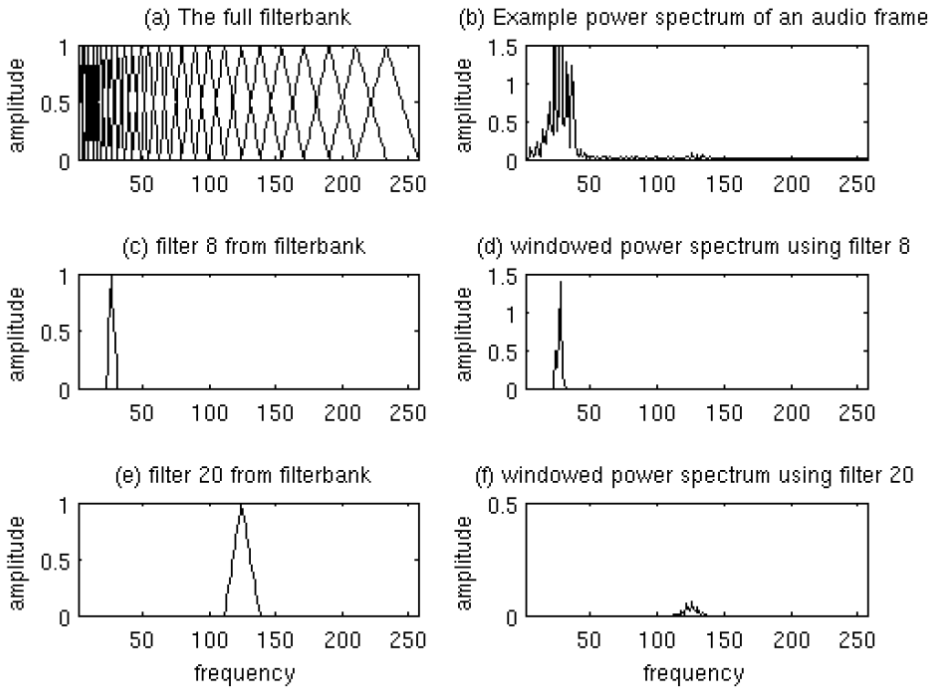

Figure 2.6 Plot of Mel Filterbank and Power Spectrum Result

Logarithm will be done on the 26 energies from previous step. Final step is to compute discrete cosine transform (DCT) of the 26 log filterbank energies. Discrete Cosine Transform (DTC) of the 26 log filterbank energies get 26 cepstral coefficients. This 26 cepstral coefficients vector is called as Mel Frequency Cepstral Coefficient (MFCC).

In speech recognition, only 12-13 of the 26 coefficients are kept. Higher DCT coefficients represent fast changes in filterbank, and this cause the degradation of the speech recognition accuracy (Ravikumar, 2009, Shrawankar, 2013).

2.3.4.3 Perceptual Linear Prediction (PLP) Cepstral

M.Hariharan et al. investigated the performance of sample entropy feature for the recognition of stuttered events. The research used perceptual linear prediction cepstral methodology for feature extraction. Three different filter banks were used to extract the sample entropy feature, which are Bark scale, Mel scale and Erb scale. The flow is shown as below:

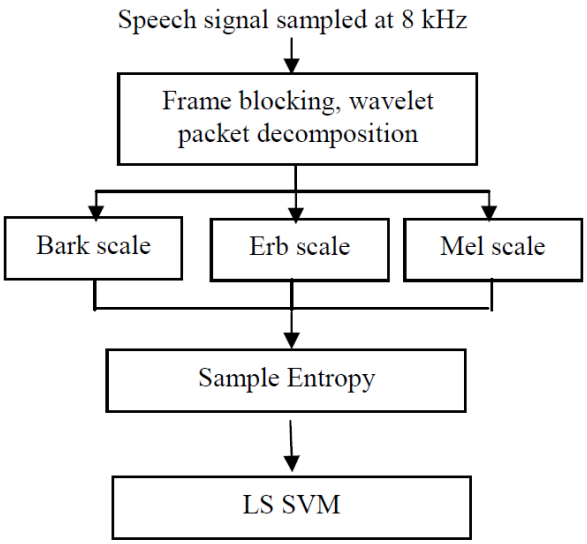

Figure 2.7 Block Diagram of Feature Extraction using Sample Entropy

Different orders of Daubechies ('dB') family wavelet packet filters was used to decompose the speech signals into various frequencies, following the Bark, Erb, and Mel scale. Sample entropy was then used to measure the complexity or regularity of a time-series signal. Pattern similarities in a time sequence can be examined, where a larger value signifies a more complex or irregular data. The characteristic is used to distinguish repetition from prolongation in the work (Hariharan et al., 2012).

2.3.5 Classifier

Classifier is used to categorize the type of dysfluent or fluent of speech. Few classify techniques are observed in the previous projects, such as:

- Artificial neural networks, ANN (Howell et al., 1997)
- Linear Discriminant Analysis classifier (LDA) and k-nearest neighbor (k-NN) (Lim et al, 2009)
- Support Vector Machine (SVM) (Ravikumar et al, 2009, Palfy et al., 2009)
- Least Square Support Vector Machine (LS SVM) (Hariharan et al., 2012)

Howell et al. had presented the recognition classifier using ANN with nine parameters, whole word and part word duration; whole word, first part and second part fragmentation; whole word, first part and second part spectral measure; and part word energy.

K-NN is a supervised learning algorithm by classifying the new instances query based on majority of k-nearest neighbor category. K-NN category is determined by a minimum distance between query instance and each of the training set calculation. Each of training instance (training speech signal) will be compared against each query instance (test speech signal). Based on majority voting of the nearest neighbor category, the k-NN prediction of the query instance is determined. The Euclidean Distance measure is used to calculate how close each member of the training set is for the test class that is being examined, which shown below (Lim et al, 2009):

$$d_E(x, y) = \sum_{i=1}^{N} \sqrt{x_i^2 - y_i^2} \quad (2.9)$$

For each test stuttered events (to be predicted), the training data set is located with k closest members (k-nearest neighbors). Class labels either repetitions or prolongations are found and class labels of test speech samples are determined by from this k nearest neighbor (Lim et al, 2009).

Linear Discriminant Analysis (LDA) is also used to categorize the speech features into 2 feature vector classes, repetitions and prolongations (Lim et al, 2009). Linear transformation (or called discriminant function) is a projection feature vector generated by transforming the speech feature vector.

Support Vector Machine (SVM) is a powerful machine learning tool which attempts to obtain a good separating hyper-plane between two classes in the higher dimensional space (Ravikumar et al, 2009, Palfy et al., 2009).

The least square support vector machine algorithm was proposed Suykens and Vandewalle in 1999. LS SVM uses a set of linear equations for training compared to the SVM which uses a quadratic optimization problem (Hariharan et al., 2012).