VOICE-BASED MALAY COMMAND RECOGNITION FOR ELECTRICAL APPLIANCES

MASYITAH BINTI ABU

FACULTY OF ENGINEERING
UNIVERSITI MALAYSIA SABAH

2017

VOICE-BASED MALAY COMMAND RECOGNITION FOR ELECTRICAL APPLIANCES

MASYITAH BINTI ABU

THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE DEGREE OF BACHELOR OF ELECTRONIC (COMPUTER) ENGINEERING

FACULTY OF ENGINEERING
UNIVERSITI MALAYSIA SABAH
2017

DECLARATION

I hereby declare that this thesis, submitted to Universiti Malaysia Sabah as partial fulfilment of the requiments for the degree of Bachelor of Electronic Engineering (Computer Engineering), has not been submitted to any other university for any degree. I also certify that the work describe herein is entirely my own, except for quotations and summaries sources which have been duly acknowledged.

This thesis may be made available within the university library and may be photocopied or loaned to other libraries for the purposes of consultation.

15 June 2017 Masyitah Binti Abu

CERTIFIED BY

Dr. Rosalyn R Porle
SUPERVISOR

CERTIFICATION

NAME	:	**MASYITAH BINTI ABU**
MATRIC NO	:	**BK13110203**
TITLE	:	**VOICE-BASED MALAY COMMAND RECOGNITION FOR ELECTRICAL APPLIANCES**
DEGREE	:	**BACHELOR OF ENGINEERING ELECTRONIC (COMPUTER) ENGINEERING WITH HONORS**
VIVA DATE	:	**8 JUNE 2017**

DECLARED BY

SUPERVISOR **SIGNATURE**

DR. ROSALYN R PORLE

ACKNOWLEDGEMENT

Assalamualaikum,

First of all, I am very thankful to Allah who is most beneficent and merciful. This thesis has been completed due to the support of many people and we wish to acknowledge them here.

I would like to express my depess gratitude and appreciation to my supervisor, Dr. Rosalyn R Porle for all her advices, guidance and support during my final year.

I also would like to express my thanks to all students Faculty of Engineering Universiti Malaysia Sabah for their kind cooperation for my thesis.

Last but not least, I would like to express my gratitude to my family for their love and encouragement. I also have received so much support from family and friends.

Masyitah binti Abu

ABSTRACT

This study is related to the Voice-Based Malay Command Recognition for electrical appliances in the living room. According to this research, there are more voice-based command recognition in English language had been build compared to voice-based command recognition in Malay language. This project is to identify the command used for electrical appliances in the living room using Malay language. In this project light and fan had been chosen as electrical appliances to build the command. All the command will be recorded in two environments: normal environment and noisy environment. The noise in the environment will be filtered in pre-processing stage. To see the performance of this project, Mel Frequency Cepstrum Coefficient will be used to extract data from the command and Neural Network will be used to analyse the command that had been chosen for voice-based command recognition for electrical appliances in the living room. The result of this project will be presented in recognition rate from each command with difference environment.

ABSTRAK

Kajian ini adalah berkaitan dengan Suara Berasaskan Melayu Pengiktirafan Arahan bagi peralatan elektrik di ruang tamu. Menurut kajian ini, terdapat lebih berasaskan suara pengiktirafan arahan dalam bahasa Inggeris telah dibina berbanding pengiktirafan arahan berasaskan suara dalam bahasa Melayu. Projek ini adalah untuk mengenal pasti arahan yang digunakan untuk peralatan elektrik di ruang tamu dengan menggunakan bahasa Melayu. Dalam projek ini lampu dan kipas telah dipilih sebagai peralatan elektrik untuk membina arahan. Semua arahan akan direkod dalam dua persekitaran: persekitaran biasa dan persekitaran yang bising. Bunyi di dalam persekitaran akan ditapis dalam peringkat pra-pemprosesan. Untuk melihat prestasi projek ini, Mel Kekerapan Cepstrum Pekali akan digunakan untuk mendapatkan data dari perintah dan Neural Network akan digunakan untuk menganalisis arahan yang telah dipilih untuk pengiktirafan arahan berasaskan suara bagi peralatan elektrik di ruang tamu. Hasil daripada projek ini akan hadir dalam kadar pengiktirafan dari setiap arahan dengan persekitaran perbezaan.

CONTENTS

CHAPTER 2 LITERATURE REVIEW

CHAPTER 3 PRE-PROCESSING COMMAND SIGNAL AND FEATURE EXTRACTION

CHAPTER 4 RECOGNITION USING NEURAL NETWORK

CHAPTER 5 CONCLUSION AND FUTURE WORK

APPENDICES

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

n	net input
i	row of matrix input
j	column of matrix input
$\propto$	learning rate
w	weight
V	performance index
k	number of layer
δ^k	Sensitivity of network
a	output of the network
b	input of the network
f	final layer of network
a_q^M	output of the network, qth is the input, p_q is presented.
e_q^T	the error for the qth input.
t	target output
x(n)	Command signal
y(n)	Output Signal

CHAPTER 1

INTRODUCTION

1.1 Introduction of Voice-Based Command Recognition

Voice is one of the gifts that had been given to human for communication between themselves. Voice is also the source of information that can be extracted and can be used for many applications. Nowadays human voice are used for the latest technologies in the field of human-computer interface, voice dialling, data entry, ID verification for money transaction, person identification in a crowd, smart home, control to access restricted area (Abhay 2015).

Furthermore, modern technology has increased the importance of communication between human and machines. In Ozturk and Unozkan (2010) research state how to improve the communication between human and machine and one of the improvement is by using voice recognition because voice has advantages such as being neutral, efficient, effective and flexible beside enable to control the machines and communicating with them. The industrial equipment or motors also can be controlled without using hands. Good voice recognition system can give many benefits especially for disabled people. Voice recognition is also suitable for wheelchairs, home appliances, system with phones and emergency systems operating by using voice-based command recognition (Ozturk & Unozkan, 2010).

Voice-based command recognition also can be used in many applications such as security devices, household appliances, cellular phones, ATMs and computer and this research is more toward electrical appliances in the living room. Nowadays, using a computer as a device for voice recognition is not very relevant because the size is big and not easy to handle, so a mobile device is one of the devices that is more suitable and easy to use (Norhafizah Aripin & Othman 2014).

Most of the mobile devices use Android for the operating system, middleware, and key applications. There are many connectivity options for Android for example Wifi, Bluetooth, and wireless data over a cellular network. The reason mobile device are more suitable compare to open air voice by using computer or microphone because there are problems such as noises, the command was given is not pronounced properly, and from the far distance, so it is hard to receive the voice command. Even usual conversations may affect the system command when it matches with the spoken word (Rahman et al. 2015).

Beside the device used, the language used for command the electrical appliances in the living room also will be discussed because there are also scientist and researchers who build computer software that can be recognized in many languages such as English, Japanese, and Thai (Rahman et al. 2015). This research use Malay language to control the electrical appliances in the living room.

By using the Malay language as voice-based command recognition, it can help ill, aged and disabled people in Malaysia who cannot speak in English. For the information, Malay language in Malaysia is used by over than 10 million speakers. The majority of Malay ethnic in Malaysia is about 54% (Malay Culture 2007) but lately, most of the research and project for voice recognition are based on English speech recognition, and in comparison Malay, speech recognition is still limited (Ong & Ahmad 2011).

1.2 Problem Statement

The main issue of voice-based Malay command recognition for electrical appliances in the living room is to help the users to control the electrical appliances in the living room using voice command. The command given for electrical appliances will be in Malay language. In the living room there are several type of background noise that may disturb the performance of the system for example the fan noise, television noise and radio noise. The noise will disturb the command signal and recognize the command given as different command.

1.3 Project Objective

The main objective of this research is to study the performance of voice based command recognition in Malay version for electrical appliances in the living room. The objective can be achieved by performing the tasks below:

a. To create Malay command database for electrical appliances in the living room.

b. To extract Malay command data using Mel Frequency Cepstrum Coefficient (MFCC).

c. To recognize the command using Neural Network Method.

1.4 Scope of Work

The scope of work is:

a. This research only focused on to turn on and turn off fan and light command.

b. There are two environments chosen which are normal environment and noisy environment to differentiate the performance of each command. For normal environment, the noise comes from the fan whereas for noisy environment, background music is used as the disturbance to imitate the sound produced form television and radio in the living room.

c. Lastly, the speaker choose for undergoes this research are multiple races in Faculty of Engineering and the age range are between 20 to 30 years old because during this age there are differences between male and female voice.

1.5 Thesis Organization

This thesis contain five chapters. All chapters are discussed as follows:

Chapter 1 elaborates the overview of voice based command recognition; Malay version and some problem statement occur in voice based command recognition system. The objective and scope of work of this research is describe in this chapter.

Chapter 2 details the literature review of this project. Data extraction was extract from some journal and paper that give enough data of this research. Comparison between journal and paper also done to justify the research gap. The review on what application used for the system, the available method used, and how to achieve the objective.

Chapter 3 discusses the methodology which covers voice based command recognition; Malay version system definition and flowchart how the system is working. How the command undergoes pre-processing and feature extraction is explained in this chapter.

Chapter 4 analyse result of voice-based Malay command recognition with the theoretical expectation on characteristics or behaviour of design as compared to a fixed decision. The operation how Voice-based Malay command recognition for electrical appliances to operate are elaborate in this chapter.

Chapter 5 describes conclusion and future work for voice based command recognition by conclude the overall of this research.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Controlling electrical appliances by using voice is one of the efforts that speech researchers want to achieve. The command given is in the Malay language, to activate the electrical appliances in the living room. Nowadays, most of the computer are based on voice recognition are in English. Due to some language barrier, the common masses of our population face big obstacle to enjoy the optimum benefits of the modern communication and information technology (ICT) where huge enriched English knowledge databases are there around the world. The only technological way that can be used to remove this barrier is language processing in mother language (Pialy et al. 2014).

In Norhafizah and Othman research, fan and light were chosen. By using voice, the speakers can control by switch on or off fan and light (Norhafizah Aripin & Othman 2014). One of the reason voice-based command recognition are done is to help handicapped people. It can help them to use electrical appliances in the living room by using simplest method. On and off command were choose because it was very suitable to control electrical appliances in the living room for example light, fan, air conditioner, television, and radio (Samah et al. 2015).

2.2 Voice Signal Acquisition and Extraction

2.2.1 Devices

For voice-based command recognition, voice is used to give the command and to record the voice command a device will be used. There is voice recognition that use microphone as a device to record the speakers' voice and when the technology become more advance some of voice recognition are using Android application or mobile device to record voice. In Norhafizah and Othman, they are using mobile device to record the voice and they said that the advantage of using mobile device is it can control the home appliances from far distances up to 20 meters (Norhafizah Aripin & Othman 2014). In Abhay research, microphone is used to record voice and the microphone used is directional Carbon Microphone. The author state that by using this microphone it can reduce the environmental noise (Abhay 2015)

2.2.2 Classification of Speech

To extract speech recognition there are many type classification of speech can be used to recognize the speech. Speech recognition had four type of classification isolated word, connected word, continuous speech and spontaneous speech. Each research had different definition for each classification, In Shikha et al. (2014) research, they state that isolated word have sample windows that accepts single utterances at a time. For connected word it is similar to isolated word but it allow separate utterance to be run together by minimum pause between them. For continuous speech, this classification allows the speaker to speak almost naturally, while the computer test the content and there are special methods used to determine utterance baundaries and various difficulties occured in the continuous speech and lastly spontaneous speech, it is a system that have an ability that can handle a variety of natural speech feature such as word being run together (Shikha et al. 2014).

In Aman et al. (2016) and Hemakumar and Punitha (2013) research, have different definition from Shikha et al. (2014) research. They state that isolated word recognizer usually require each utterance so it can have quiet on both sides of the sample window. Single word usually is not accept but it requires a single utterance at a time for example one word responses or commands but for multiple word inputs it is very unnatural. The major advantage of using this type of classification, it is easier to implement and comparatively simpler because the word boundaries are available and the word tent to be clearly pronounced and the disadvantage of this application is it choose different boundaries that will effected the result. The second classification is connected word, it has similar definition as Shikha et al. (2014) research. The third classification continuous speech, this classification allow speaker to speech almost naturally and the computer will determines the content while the speaker is speak. Adjacent word also can be run together without pause or any other apparent division between words. The disadvantage using continuous speech is as the vocabulary size grow larger, confussion between words also grow. For spontanous speech, it is not rehearsed but natural and this type of classification should be able to handle many type of natural speech feature such as word being run together and even slight stutter. The disadvantage of this classification is it maybe include the mispronunciations, false-start and non-word (Aman et al. 2016; Hemakumar & Punitha 2013).

2.2.3 Speech Characteristic

There also several speech recognition techniques that had been used for determined speech characteristic because each speech data contain different type of information based on the speaker vocal tract, excitation source and behaviour feature. There are three types of speech analysis used, first analysis is segmentation analysis, in this technique speech will be analysed using frame size and shift in the range of 10-30 ms to extract speaker information. This method is used to extract vocal tract information of speaker recognition. Another technique is sub segmental analysis and it also using he frame size for speech analysed but it only shift in range of 3-5 ms. This technique is used to analyse and extract the characteristic of the excitation state. The last technique is supra segmental analysis, it using frame size also and this technique is used to analysed and characteristic the behaviour character of the speaker (Shikha et al. 2014; Aman et al. 2016; Om & Navneet 2013).

2.2.4 Pre-processing Technique

There are some techniques used to get the command performances. Firstly, the speaker will give their command to activate the living room electrical appliance in Malay language. The natural voice of the speaker will be identified and transformed into a voice signals through recording. The voice signal will be added to input of recognition system for pre-processing. In Wenhao et al. (2010) research, they includes voice signal sampling, anti-aliasing band pass filtration, removing individual differences in pronunciation and noise caused by environment in it pre-processing process and endpoint detection and element selection of speech recognition will be involved (Wenhao et al. 2010). While in Jun & Xiaoxiao (2011) research, they send the recorded voice toward voice activation detection to detect the strat point of people speech and eliminate the useless noise and signal. Then, pre-empt will be done by improving the high frequency segment to balance the decreasing in the spectrum analysing. After that, framing will be done to dividing the digital input to specifics frames. Last step in the pre-processing is low-pass and high-pass filtering, low-pass and high-pass filtering is to filtering the high and low frequency noise (Jun & Xiaoxiao 2011).

2.2.5 Feature Extraction Technique

Here there are three feature techniques that can be used to extract the voice, feature extraction will extract the command signal into a sequence of vectors. Yoghesh and Mukta (2015) explain about three type of feature one of it is MFCC that had been explain before and the other feature is Linear Predictive coding (LPC) and Perceptual Linear Prediction (PLP). LPC is a mathematical computational operation that linear combination of several previous samples. LPC of speech has become the predominant technique for estimating the basic parameters of speech. It provides both an accurate estimate of the speech parameters and it also an efficient computational model of speech. The basic idea behind LPC is that a speech sample can be approximated as a linear combination of past speech samples. Through minimizing the sum of squared differences over a finite interval between the actual speech samples and predicted values, a unique set of parameters or predictor coefficients can be determined. These coefficients form the basis for LPC of speech. PLP model developed by Herman sky (1990). The goal of the original PLP model is to describe the psychophysics of human hearing more accurately in the feature extraction process. PLP is similar to LPC analysis, based on the short-term spectrum of speech. In contrast to pure linear predictive analysis of speech, perceptual linear prediction (PLP) modifies the short-term spectrum of the speech by several psychophysically based transformations (Yoghesh & Mukta 2015).

2.3 Recognition Methods

For voice-based command recognition, there are many methods used to recognize the command. There are several types of methods for voice-based recognition; neural network method (NN), fuzzy logic technique, hidden Markov model (HMM), and Gaussian mixture model (GMM) (Ong & Ahmad 2011). But for this research, it will focus more at HMM and ANN method because this method is frequently used for voice-based command recognition. For ANN methods, it can be used for estimating posterior probabilities and training the network and HMM method can be used for decoding and language modelling.

Jianliang et al. (2012) combine two methods to improve the voice recognition, first it state that a more complete expression of acoustic model of the voice are suitable to use HMM and Statistical this methods of training the underlying acoustic model. This research also mentions that the upper voice model into the unified voice recognition search algorithm also had been used by HMM so it can obtain better recognition result. The other method are artificial neural network, it state that this method is applied to solve the pattern classification problems and was shown to have enormous energy. Speech recognition neural network are usually divide into two categories, a class of neural network or neural network with traditional HMM (hybrid HMM/ANN) and the other one is the establishment of auditorial neural network model based on human auditory physiology (Jianliang M. et al. 2012).

2.3.1 Hidden Markov Models

Hidden Markov models (HMM) has widely used statistical approach to characterize the spectral properties of frames of speech in the context of statistical methods for speech recognition. HMMs have also an advantage of providing a natural and highly reliable way of recognizing speech for a wide variety of applications as a stochastic modelling tool since HMM integrates well into systems incorporating information about both acoustics and syntax. It is currently the predominant approach for speech recognition. Hidden Markov Models are "doubly stochastic process" in which the observed data are viewed as the result of having passed the true process that is called hidden process through a function that produces the second process that call observed process. The hidden process consists of a collection of states connected by transitions, each transition is described by two sets of probabilities. First probability is called transition probability, which provides the probability of making a transition from one state to another. Second is an output probability density function, which defines the conditional probability of observing by using the combination of following algorithms the HMM is implemented by three algorithm Forward-backward algorithm, Viterbi algorithm and Baumwelch algorithm (Urmila & Anjali 2013).

Anand et al. (2014) had made a research about voice recognition based on HMM. This research mentions that HMM is a statistical modelling approach and is defined by initial state probability, the state transition probability matrix, and the output probability matrix. The computation cost is very high for the typical HMM-based speech recognition algorithm. It is depends on the number of state for each word and all words, the number of Gaussian mixture, the number of speech frames, number of feature for each speech frame and lastly the size of vocabulary (Anand et al. 2014).

2.3.2 Artificial Neural Network

Besides that, there other methods that also being used in voice based command recognition is the method that called Artificial Neural Networks (ANN) or Neural networks. In this method, it often used as a powerful discriminating classifier for tasks in automatic speech recognition and they have several advantages over parametric classifiers. However there are disadvantages in terms of amount of training data required and length of training time. There are some neural network architectures for example Feedforward Perceptrons Trained with BackPropagation, Radial Basis Function (RBF) Networks and Learning Vector Quantization (LVQ). Based on three architectures, Feedforward Perceptrons Trained with BackPropagation is most popular model in Neural Network training. However, the length of time required to train the networks can present problems, particularly when investigating a variety of feature sets to represent speech data (Urmila & Anjali 2013).

Pialy et al. (2014) research, state this speech recognition technique it follow a template matching of features with time normalization by dynamic time warping or it also can be called neural network model. There is a feature used to create template, it is typical drive from spectral or log spectral representation of each frame of speech and the feature that commonly used is parameter from linear prediction model and mel-cepstral coefficient. This research uses neural network model as their modelling background. Firstly, they record the command from the speaker to be recognize in computer through MATLAB. From the recorded voice Mel-frequency cepstrum coefficient extract after several mathematical computation. Then the acoustic voice signal will be converted to set the numerical values. Using Mel-frequency cepstrum coefficient feature, training dataset is create based on the speaker and the referred voice. Besides that, target data table also being created as backpropagation neural network. Lastly, the test data is compared with the train dataset (Pialy et al. 2014).

2.4 Performance Criteria

For this research, the technology used are electrical appliances in the living room. Here there are two researches that used three different environments for their experiment; noiseless, less noise and noisy. For first research, its recognition rates of the experiment are 98.2% in a quiet room, 93.0% in an office room and 44.2% in a noisy room and for the other research, its recognition rates for the noisiest environment was 49.2%, medium noise level is at 63.8% and lastly for quiet background is at 75.7%. In conclusion, as the noise become highers it is hard to recognize the speech because the speech signal will be interfered with the noise signal (Parichart et al. 2010; Rawan et al. 2012).

For the feature extraction, each feature gives different recognition rate. Yoghesh and Mukta (2015) compare the feature extraction recognition rate for three signal: for first signal MFCC 51.25% LPC 37.5% PLP 49.5%, for second signal MFCC 86.67% LPC 80.5% PLP 77.4%, for third signal MFCC 93.6% LPC 76.6% PLP 90.4% and for the last signal, MFCC 96.5% LPC 65.8%PLP 78.5% (Yoghesh & Mukta 2015). Poonam and Anjali (2016) also compared percentage between different feature extractions. They only used one signal and different types of features for the comparison Mel-Frequency Cepstral Coefficient (MFCC) is 78%, Perceptual Linear Predictive (PLP) is 60% and for Mel-Frequency Cepstral Coefficient – Perceptual Linear Predictive (MFCC-PLP) as a feature compare to LPC it recognition rate is 79% (Poonam & Anjali 2016).

In Poonam and Anjali also tell about the accuracy performance for method use, they tell that an average accuracy of 78% was found for male speaker whereas 80% for the female speaker. Therefore, 79% of the average recognition result is obtained as an output of the algorithm. Neural Network, had three technique: Feed-Forward BPN, Perceptron Neural Network and Linear Neural Network, it state that the average accuracy of this three technique is 79% for the feedback, 73% for perception neural network, and 69% linear neural network (Poonam & Anjali 2016). Pialy et al. (2014) said when ANN is used for speaker recognition, the recognition rate is very close to 83%, where for speech recognition exactly 60% (Pialy et al. 2014). For other HMM, the recognition rate is relatively slights higher than the average one, which is about 75%

for the noisy environment. Admittedly, however, the environment we use is not under the low SNR. For female tester, the recognition rate is 100% (Jun & Xiaoxiao 2011).

2.5 Summary

In this chapter, after comparing several recognition techniques and how voice signal are acquisition and extract some pros and cons in each of the techniques and how the signal is extract and acquisition. Word classification for voice-based Malay command recognition for electrical appliances is continuous speech because it will be given a command toward electrical appliance such as 'Buka Kipas' and the device use to recorded the voice is mobile phone because it can be record at any distance.

To test the command performance two environments will be choose: normal environment and noisy environment to see the differences. Artificial Neural Network is used in this research because it is easy to understand compare to Hidden Markov Method. For feature extraction, MFCC is used because recognition rate for MFCC is higher compare to LPC and PLP.

CHAPTER 3

PRE-PROCESSING COMMAND SIGNAL AND FEATURE EXTRACTION

3.1 Introduction

The aim of this research is to study the performance of voice-based command recognition in the different backgrounds by using Neural Network Method. The speakers need to record their voices for three times. For every speakers, one out of the three samples of the voice is used as the training data, and the remaining two samples are tested to observe the performance in recognition rate. The command signal undergoes pre-processing and feature extraction to improve the signal performance. Then, to recognize the command signal Neural Network method is used to train the training data. The test data is test with training data for pattern matching. Figure 3.1 show the flowchart of voice-based Malay command recognition.

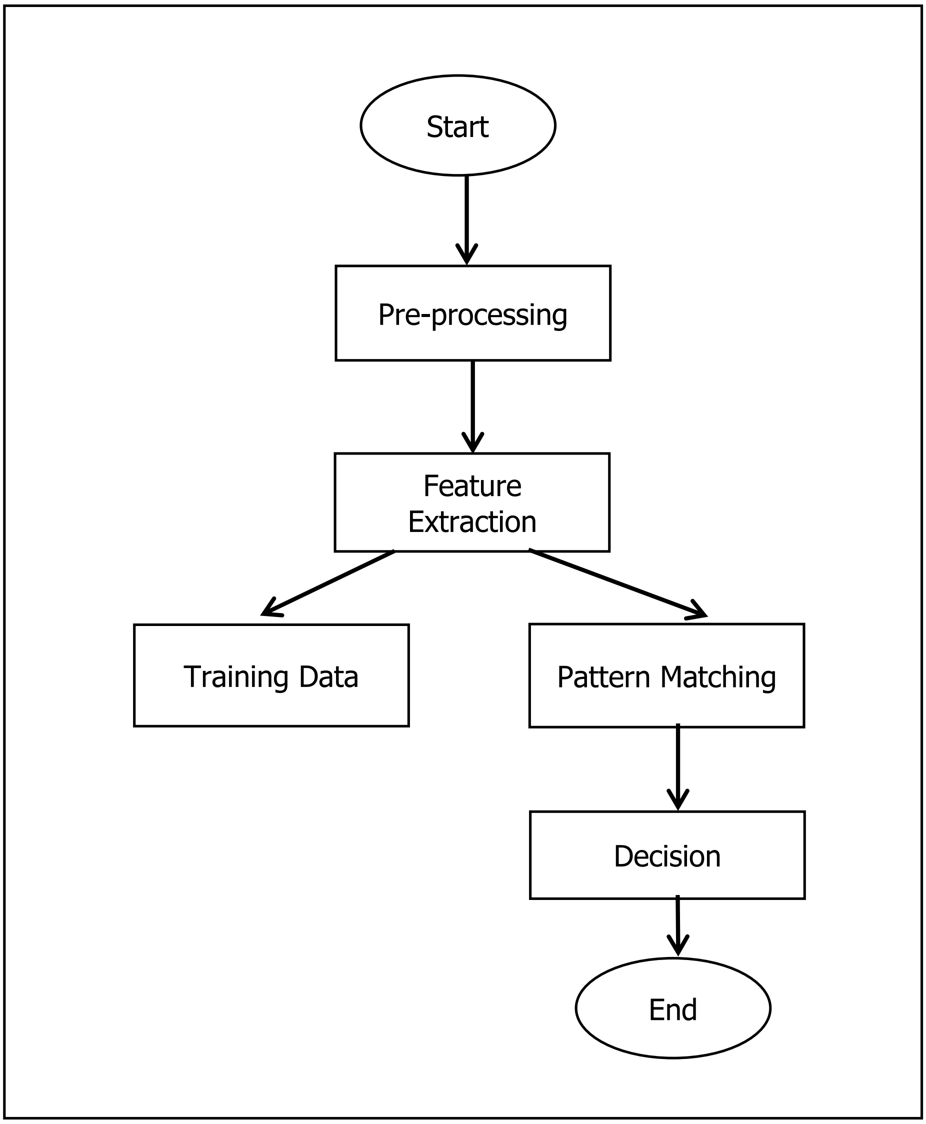

Figure 3.1 Flowchart of Voice-based Malay Command Recognition

3.2 Malay Command Database

For Malay command database, four commands are produce from two chosen electrical appliances. The electrical appliances chosen are fan and light so the commands are 'Buka Kipas', 'Tutup Kipas', 'Buka Lampu' and 'Tutup Lampu'. Table 3.1 shows the Malay commands for electrical appliances.

Table 3.1 Malay Commands for Electrical Appliances

Devices	Malay Command
Fan	Buka kipas Tutup kipas
Light	Buka lampu Tutup lampu

In the database, there are 720 samples. There are four command and each command has 180 samples. The command is recorded in normal and noisy environment so each environment has 90 samples. From the samples, 120 samples are used as training data so the samples is train in Neural Network. 75% from the samples is used for training, 10% from the samples is used for validation and 15% from the samples is used for testing result. The speakers with the range of age 20 to 30 years old are asked to record their voice. The speakers voice is record in class room because it has same environment as living room.

The voices are record using a phone recorder with same distance. The distance between the speaker and phone is about 0.5 meter and the time range to recorded speakers' voices is approximately 2 second. After recording, the voice will be converted into wave format using online software (online-audio-converter.com). The sampling rate is set to 8 kHz and the recording channel use is mono.

The samples from Male 1 is chosen to show the command signal of each command. All commands produce four signals pattern because they have four phonemes. For 'Buka Kipas' command, 'bu' is the first phoneme. The amplitude for this phoneme is between 0.01 and -0.01. The second pattern is 'ka', the amplitude for this phoneme is between 0.04 and -0.04 and the time range for this phoneme is between 0.25 s and 0.4 s. For the third pattern is 'ki', this phoneme amplitude is between 0.02 to -0.02. Last pattern for this command is 'pas', this phoneme has the longest time range between 0.6 s until 0.95 s and the highest amplitude is between 0.04 and -0.04 and decrease proportionally until it reach the lowest amplitude between 0.01 and -0.01. Figure 3.2 shows the signal of 'Buka Kipas'.

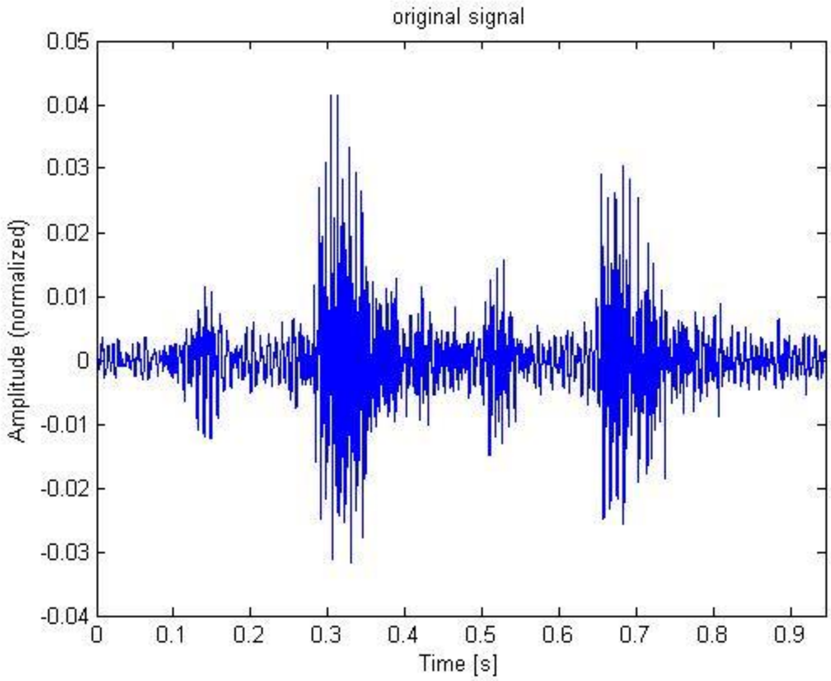

Figure 3.2 Command signal 'Buka Kipas' for Male 1 in a Normal Environment

For 'Tutup Kipas' command, the first pattern is 'tu' phoneme. This phoneme has amplitude between 0.02 and -0.02 it has higher amplitude compare to 'bu' phoneme. For second pattern is 'tup' phoneme, this phoneme has same amplitude as 'ka' but different time range. The time range for 'tup' phoneme is longer than 'ka' phoneme between 0.2 s until 0.4 s and it has higher amplitude at start signal and low amplitude at the end of the signal. For 'ki' and 'pas' phonemes is explained in 'Buka Kipas' command. Figure 3.3 shows the command signal of 'Tutup Kipas'.

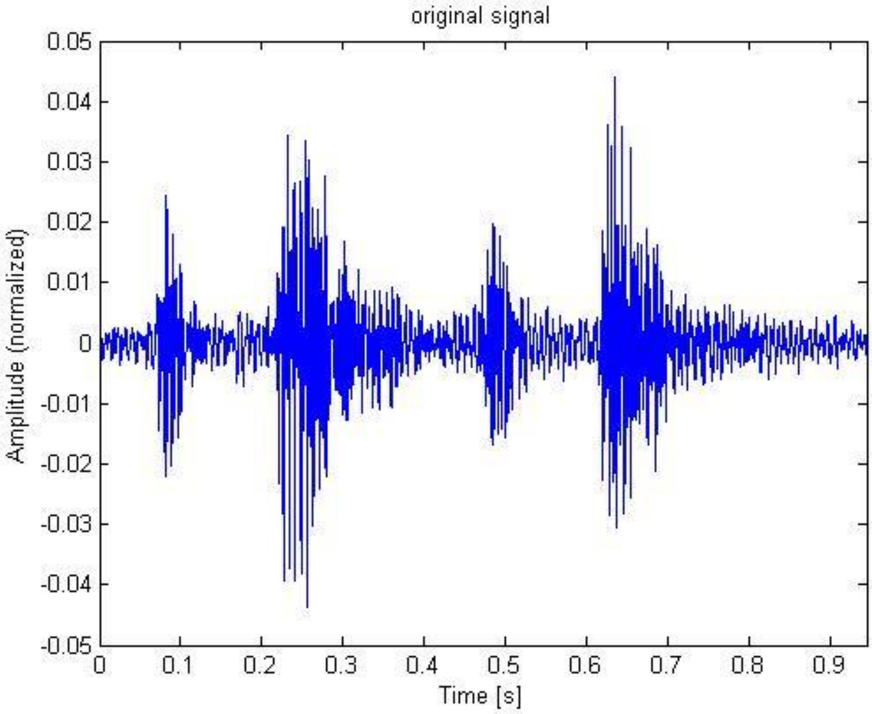

Figure 3.3 Command signal 'Tutup Kipas' for Male 1 in a Normal Environment

For 'Buka Lampu' command, the first pattern is 'bu' and second pattern is 'ka' same as first and second pattern in 'Buka Kipas'. For third pattern is 'Lam' phoneme, this phoneme has time range is between 0.4 s until 0.7 s. The highest amplitude of this phoneme is between 0.04 and -0.06. The pattern of this phoneme is increase from the lowest amplitude to highest and decrease proportionally until the lowest amplitude. The lowest amplitude of this phoneme is between 0.01 and -0.01. For 'Buka Lampu' signal the pattern for 'ka' and 'lam' phonemes is connected. For 'pu' phoneme, the pattern is the same as 'bu' phoneme. Figure 3.4 shows the command signal of 'Buka Lampu'.

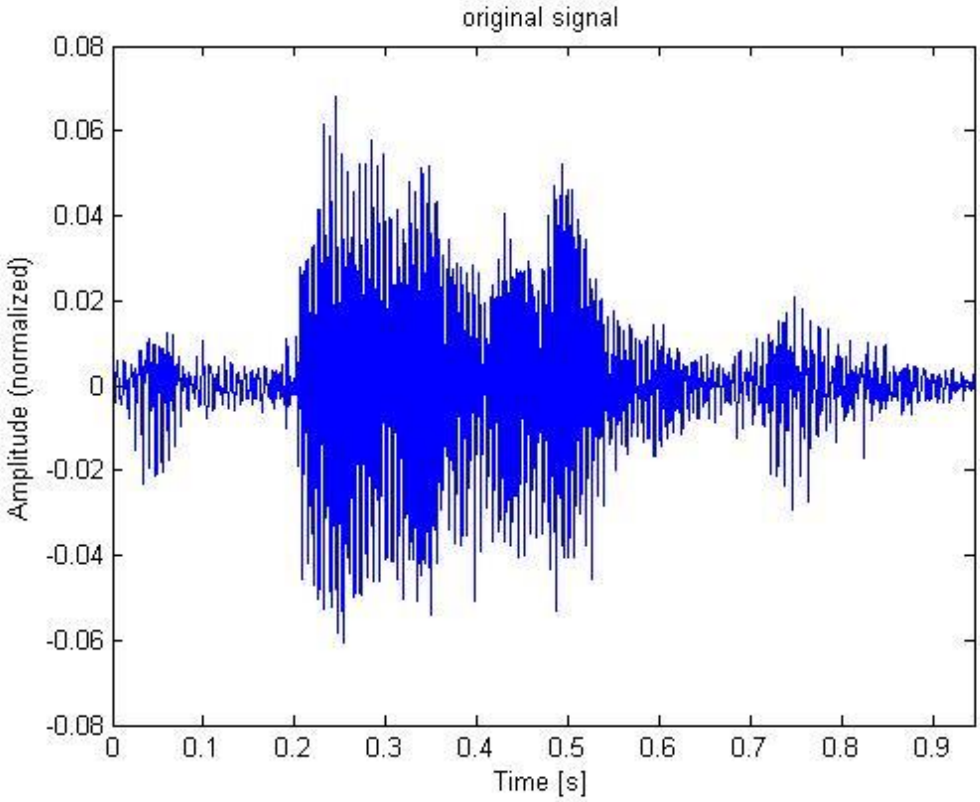

Figure 3.4 Command signal 'Buka Lampu' for Male 1 in a Normal Environment

Lastly, for "Tutup Lampu" command all the phonemes had been explained from other command. The difference in 'Tutup Lampu' command is the 'tup' and 'lam' phonemes is not connected and most of the phonemes pattern has longer time range compare to other command signal. Figure 3.5 shows command signal of 'Tutup Lampu'.

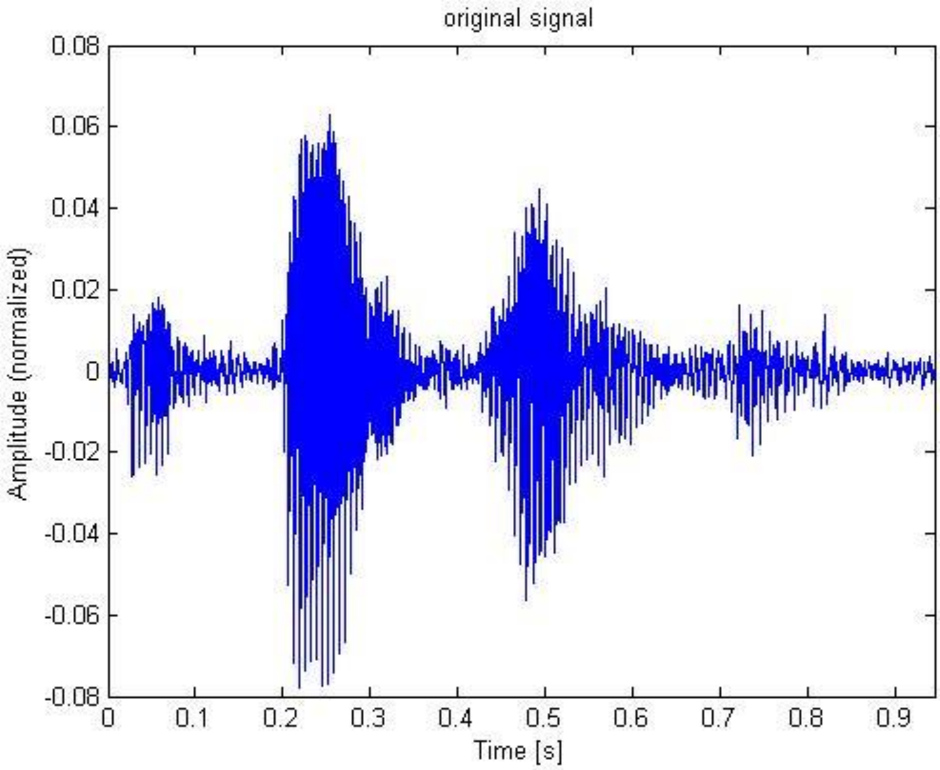

Figure 3.5 Command signal 'Tutup Lampu' for Male 1 in a Normal Environment

The command pattern for noisy environment also have the same command pattern as normal environment. The different is only the amplitude of command signal. The amplitude of command signal for noisy environment is higher than normal environment because the louder background environment, the louder the speakers will speak. Figure 3.6 shows command signal of 'Buka Kipas' in a noisy environment.

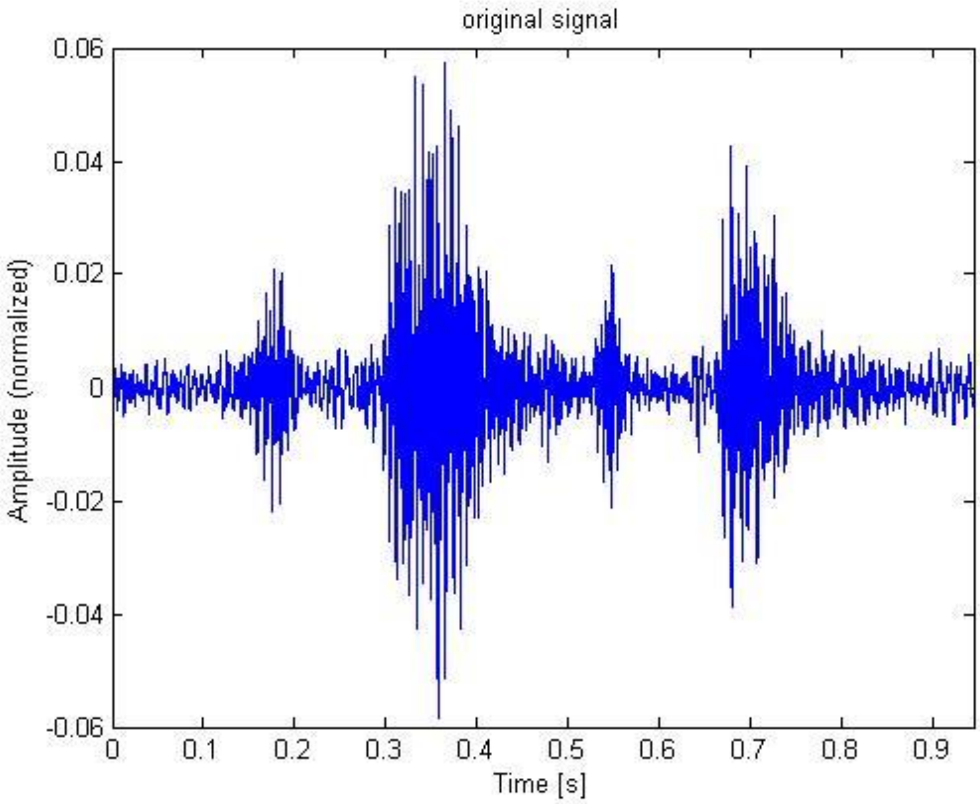

Figure 3.6 Command signal 'Buka Kipas' for Male 1 in a Noisy Environment

After the command signal is recorded, it is processed in the pre-processing stage.

3.3 Pre-processing

For pre-processing, the first step is to apply normalization of time domain for command signal to make sure all command signal time are same. Firstly, endpoint and frame shifting is used to normalize the time domain for command signal but it gives different time domain for each speech. Lastly, the command signal is edited manually by audacity. The command signal time is edit 0.95 s second for all the speech.

After that, filtering is applied to remove noise. To improve high frequency segment in the command use pre-emphasis filter to balance the decreasing in spectrum analysis. Pre-emphasis is a first order FIR filter. The transfer function is Equation 3.1: (Jun & Xiaoxiao 2011). By doing pre-emphasis the amplitude of command signal amplitude decrease and the voice become high pitch.

$$y(n)=x(n)-x(n-1) \tag{3.1}$$

Where

$x(n)$ = is the command signal

$y(n)$ = is the output signal

After pre-emphasis filter, the command signal will be filtered using low pass or high pass filtering. Low pass filter is done to remove the hissing sound in the command signal but high pass filter cannot be use to filter the command signal because it can only filter if the cut-off frequencies within the interval of 0 and 1, most of the command signal recorded are not within the interval. This low pass filter also make the voice loud. The filter used is Butterworth filter. Figure 3.7 shows the signal after pre-emphasis and low pass filtering and original command signal 'Buka Kipas'. From the figure, for low filter the noise signal is removed from the command signal and the command signal become more clear compare to original signal.

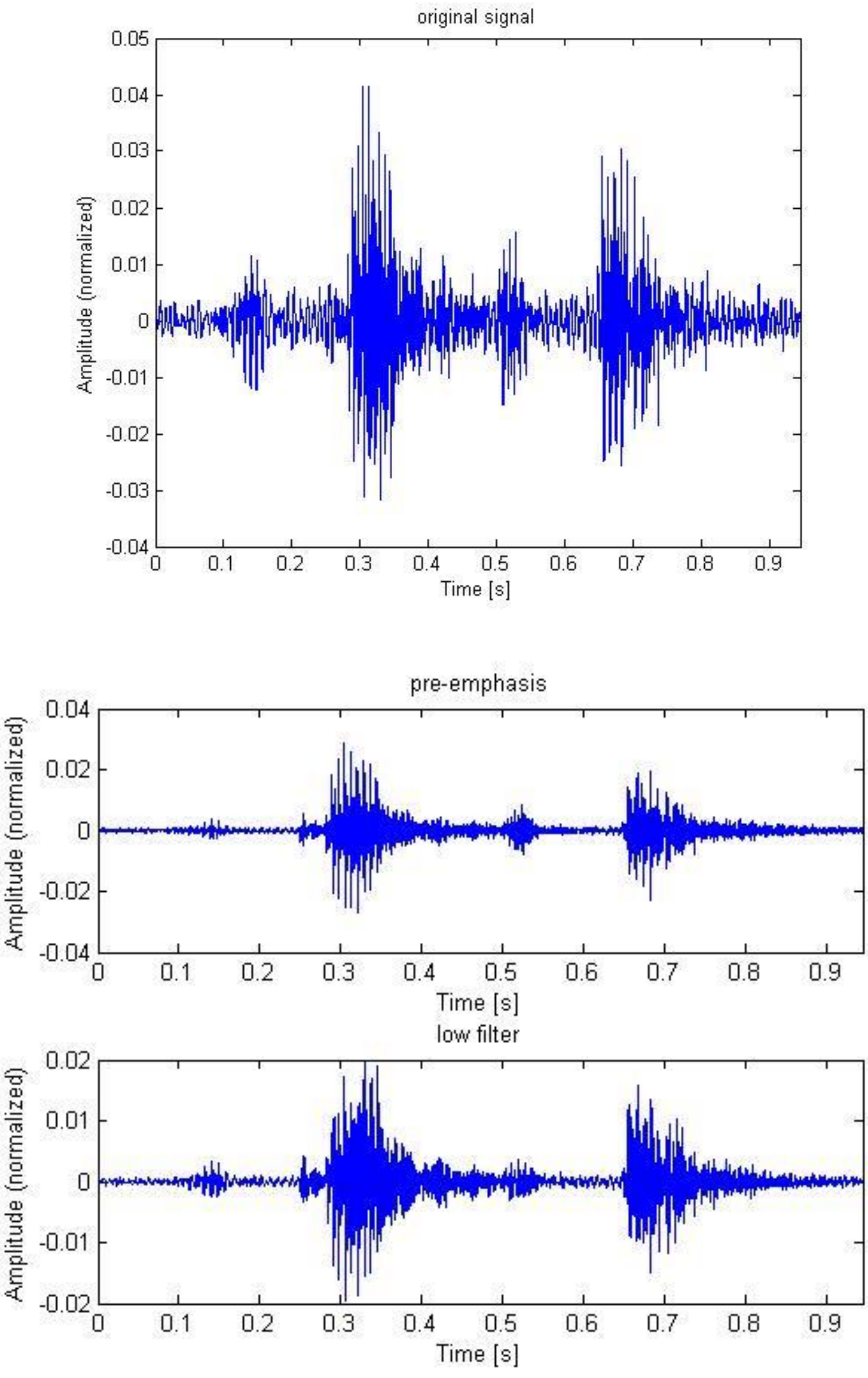

Figure 3.7 Original Command signal, pre-emphasis and low pass filtering Command signal 'Buka Kipas' for Male 1 in a Normal Environment

Beside Butterworth's low pass filter and high pass filter, other filters also is tested to remove noise from command signal for example low pass IIR, low pass FIR and bandstop IIR. Low pass IIR is Butterworth's low pass filter because the affect show by both filters are the same, it make the command become louder but it not remove the hissing sound. For low pass FIR the voice produce by this filter is to small and cannot be heard and for bandstop IIR it produce the same sound as original sound but slightly high pitch because it filters all the command. The differences for the three filters is the sound produce by the filter. Figure 3.8 shows the filter output. The blue colour is original signal and the green colour is the filtering signal.

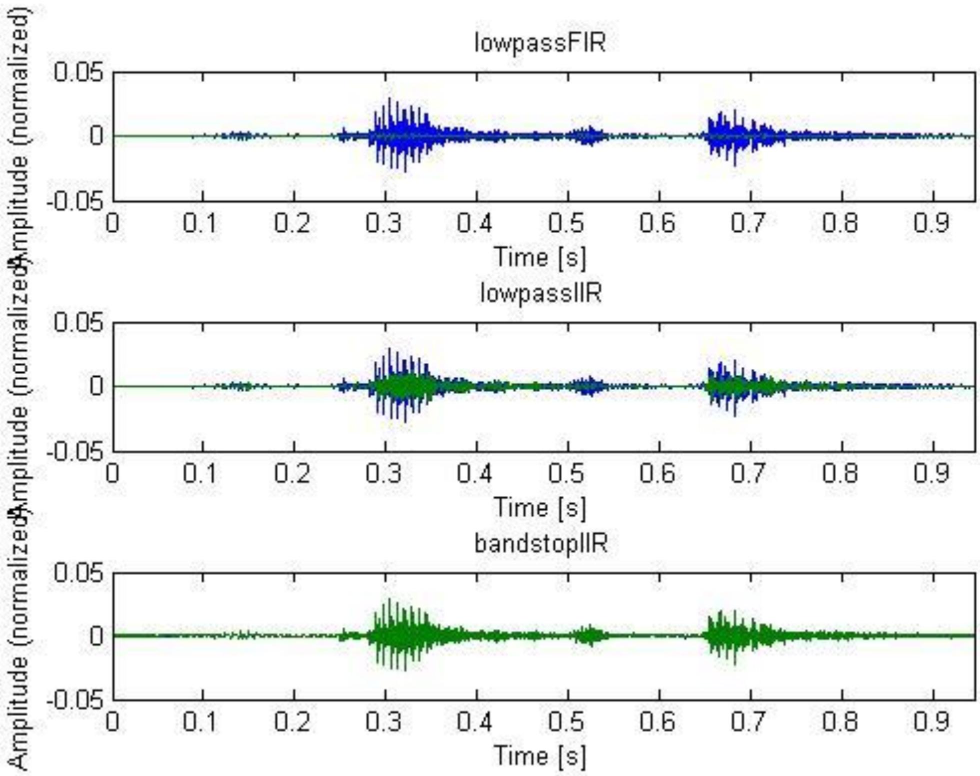

Figure 3.8 Not Suitable Filtering for Command signal 'Buka Kipas' for Male 1 in a Normal Environment

After the command signal is pre-processing, the command signal is extracted using Mel Frequency Cepstrum Coefficient in the feature extraction stage to get dataset of the command signal.

3.4 Feature Extraction

In feature extraction, Mel-Frequency Cepstrum Coefficient (MFCC) is used because it has higher recognition rate compared to other feature extraction. MFCC feature will extract represent spectral information from the command signal (Fadhilah Rosdi 2008). To extract the dataset from the command signal, there is some step need to follow.

The first step for this method is framing and windowing. In this step block framing is used because block framing crop the command signal to remove silence or acoustic interference at the beginning or end of the input command signal (Pialy et al. 2014). For the framing overlapping, the frames are 50% overlapping to remove the cut-off effect (Jun & Xiaoxiao 2011). For windowing Hamming window is applied to each frame to minimize the discontinuities of the signal by tapering the beginning and end of each frame to zero. The command signal is divided into 29 frames. The coefficient of a Hamming window is computed Equation 3.2. Figure 3.9 shows framing and windowing for the command signal 'Buka Kipas'.

$$w(n)=0.54+0.46\cos\left(2\pi\frac{n}{N}\right),0\leq n\leq N \qquad (3.2)$$

Where

N = is frame length

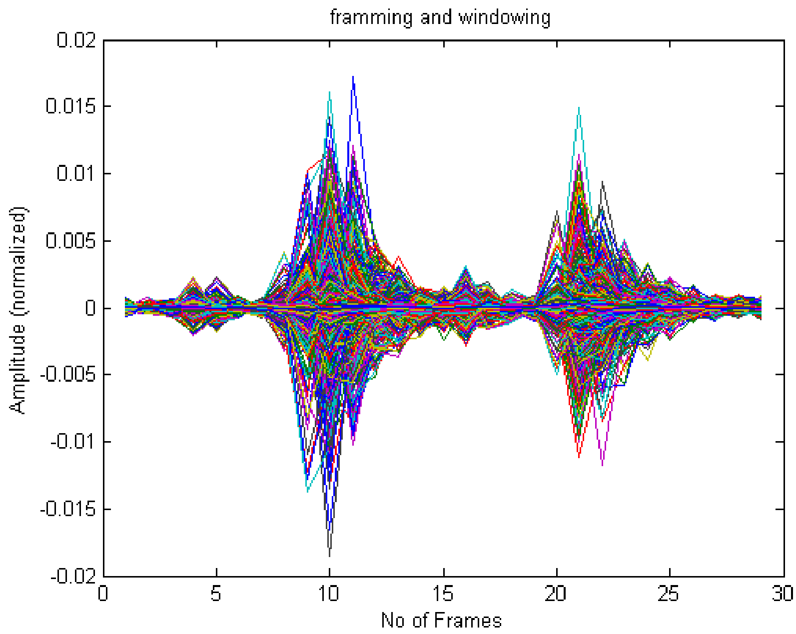

Figure 3.9 Framing and Windowing Command signal 'Buka Kipas' for Male 1 in a Normal Environment

After that, each frame needs to be Fast Fourier Transform (FFT) to change the command signal into power spectrum. The signal describes the frequency content of the signal over time. The FFT is used for increase efficiency due to rapid transformation ability. Then, the power spectrum that obtains from FFT is wrapped according to Mel Scale. Mel Frequency Scale is linear frequency spacing below 1000 Hz and a logarithmic above 1000 Hz. Therefore, as a nonlinear transformation of frequency scale, Equation 3.3 is used (Pialy et al. 2014). Figure 3.10 shows the power spectrum of the command signal.

$$melmax = 2595*\log10(1+fmax/700) \tag{3.3}$$

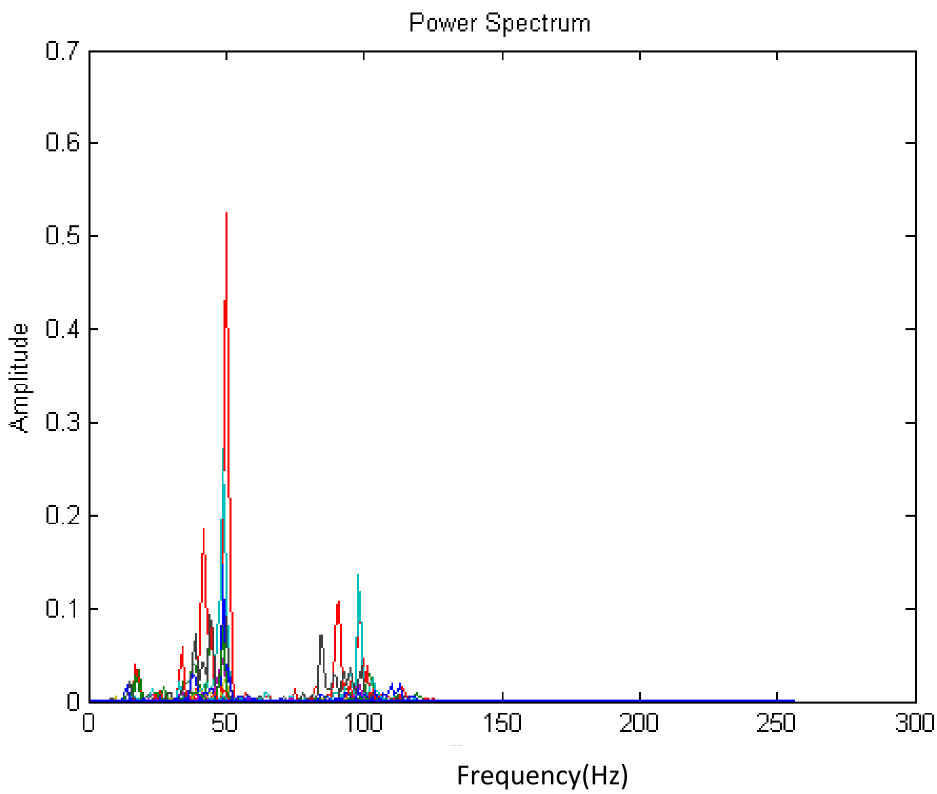

Figure 3.10 Power Spectrum Signal 'Buka Kipas' for Male 1 in a Normal Environment

Lastly, log Mel Spectrum will convert back to time domain and the result obtained will be called Mel Frequency Cepstrum Coefficient (MFCC). In the cepstrum analysis, band pass filter is applied to command signal spectrum. The filters have centre frequencies equally spaced to Mel values and it corresponds to different frequency value in Equation 3.4. In this process, only 20 number of channel is applied instead of 40 because the command signal already filter in pre-processing stage. The band pass filter is also normalized to make sure all the command signal have the same size. Figure 3.11 shows the signal of band pass filter.

$$\text{fcenters} = 700*((10.\hat{}(\text{melcenters}./2595))-1) \quad (3.4)$$

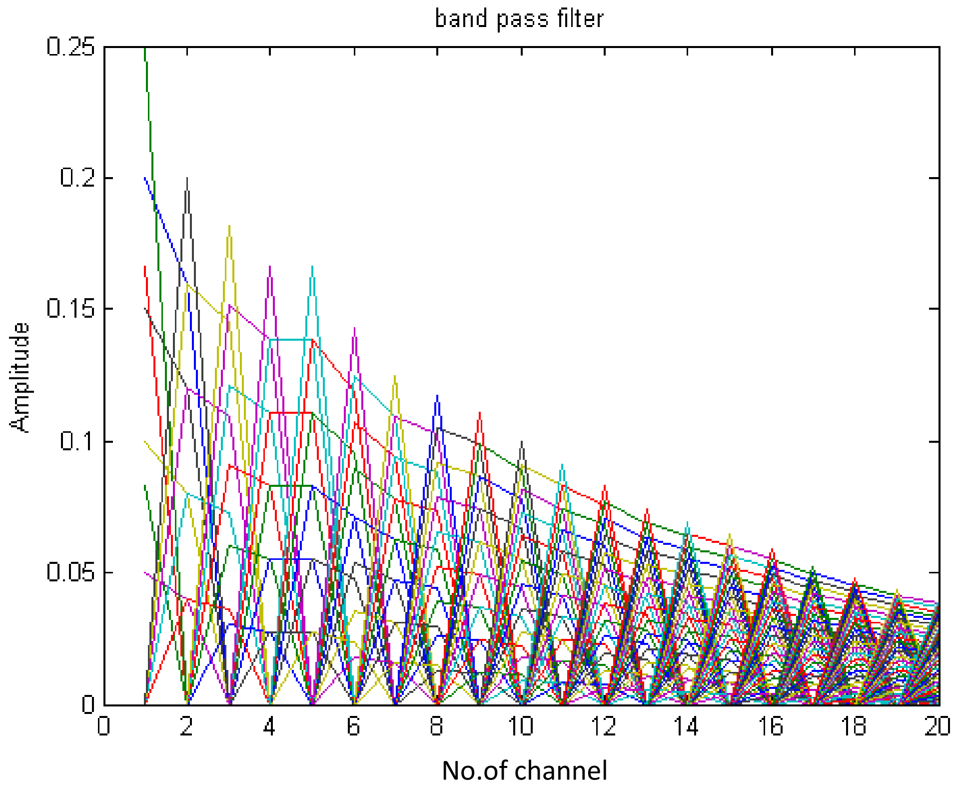

Figure 3.11 Band pass filtering of Command signal 'Buka Kipas' for Male 1 in a Normal Environment

The Mel spectrum coefficient is in real number so the command signal is converted to time domain again using the discrete cosine transform (DCT). Finally, the MFCC feature of the command signal is obtain. After done all the process for feature extraction, the MFCC feature obtain is 19 x 29 feature for each command signal. For 'Buka Kipas' command, it MFCC feature is compact compare to other commands. Figure 3.12, shows MFCC feature of 'Buka Kipas'.

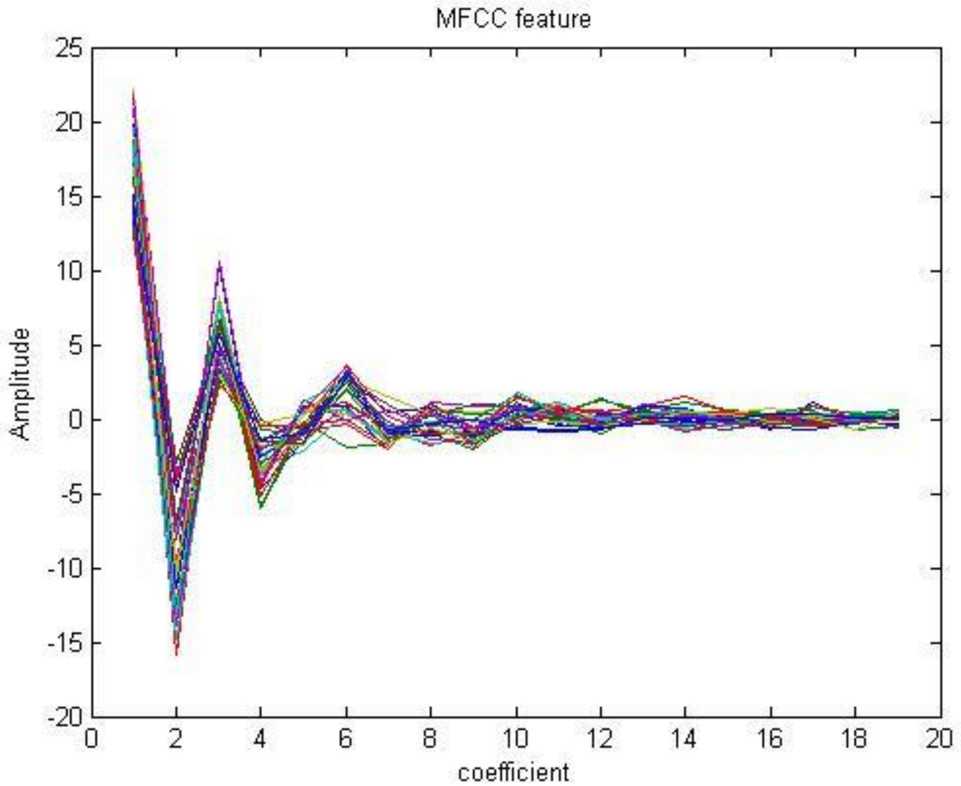

Figure 3.12 MFCC of 'Buka Kipas' for Male 1 in a Normal Environment

For 'Tutup Kipas' MFCC feature, it has the same pattern as 'Buka Kipas' but at coefficient 4 until 6 the framing is lost compare to MFCC feature in 'Buka Kipas'. Figure 3.13 shows MFCC feature of 'Tutup Kipas'.

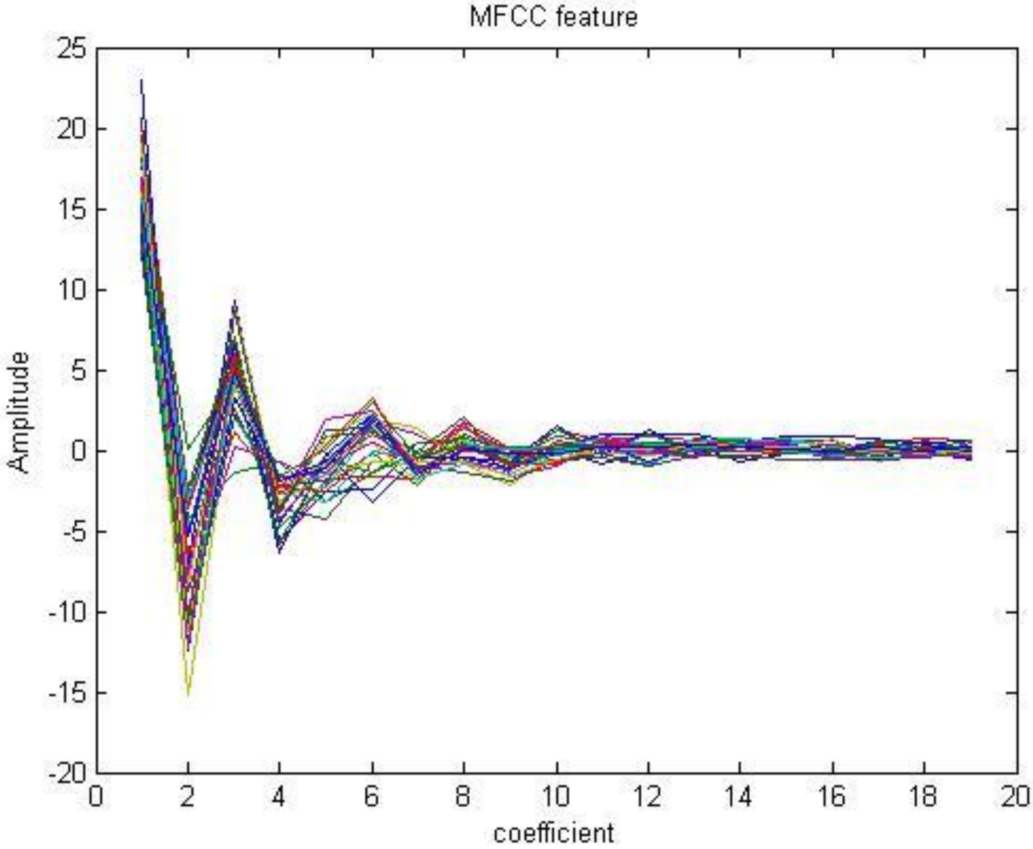

Figure 3.13 MFCC of 'Tutup Kipas' for Male 1 in a Normal Environment

For the third command 'Buka Lampu', it MFCC feature pattern is slightly different from 'Buka Kipas' and 'Tutup Kipas' MFCC feature. It only has four high peaks compare to other two commands that has five high peaks. The pattern for third peak also not as slope as the other two command. Figure 3.14 shows MFCC feature of 'Buka Lampu'.

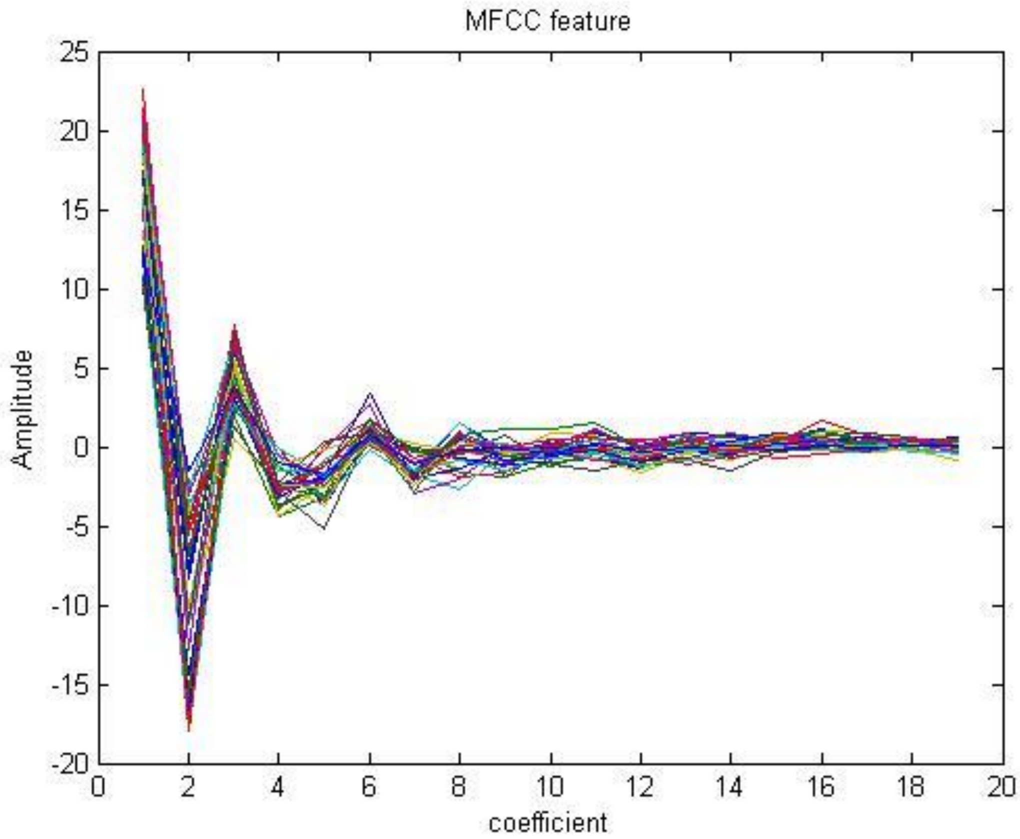

Figure 3.14 MFCC of 'Buka Lampu' for Male 1 in a Normal Environment

For the last command signal 'Tutup Kipas', the feature is not compact as other command MFCC feature but it has the same pattern as 'Buka Lampu' MFCC feature pattern. The framing at coefficient 4 until 8 is loose compare to 'Buka Lampu' MFCC feature. Figure 3.15 shows MFCC feature of 'Tutup Lampu'.

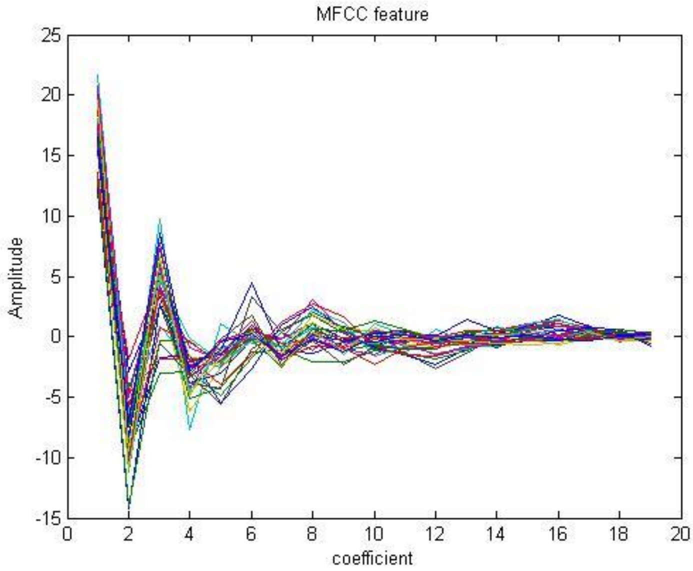

Figure 3.15 MFCC of 'Tutup Lampu' for Male 1 in a Normal Environment

The MFCC feature pattern for noisy environment also have the same pattern as normal environment because the command signal is undergoes filtering to remove noise. Figure 3.16 shows command signal of 'Buka Kipas' in anoisy environment.

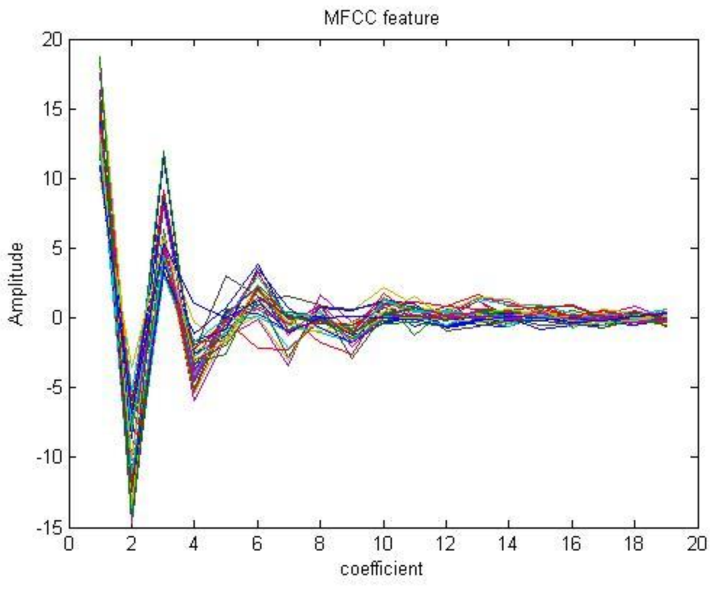

Figure 3.16 MFCC of 'Buka Kipas' for Male 1 in a Noisy Environment

3.5 Summary

This chapter discussed the overall Voice-based Malay command recognition for electrical appliance flow. Firstly, this chapter explain how Malay command of voice-based Malay command recognition is chosen based on electrical appliances used. The Malay command that had been chosen for this chapter are 'Buka Kipas', 'Tutup Kipas', 'Buka Lampu' and 'Tutup Lampu'. Then, 720 samples are recorded each command has 180 samples. Each command is recorded in normal and noisy environment so each environment has 90 samples. From 720 samples, 120 samples are used as reference model. This chapter also it also discusses software implementation for this research. Besides that, the sequence on how to process the signal also had been discussing in this chapter.

The detailed sequence on how to process the signal had been discussing based on the recognition method used. First, the input the voice command signal and undergo pre-processing such as extracting the characteristic of the signal. After that, feature extraction will occur based on the feature choose. Then, the input will undergoes pattern matching based on training data. Lastly, follow the judging rules to produce a recognition result.

CHAPTER 4

RECOGNITION USING NEURAL NETWORK

4.1 Introduction

The performance of voice-based Malay command recognition is analysed using Neural Network method. Therefore, the dataset of training data that extract by Mel Frequency Cepstrum Coefficient (MFCC) is used as an input for Neural Network. There are 120 samples of training data so each command has 30 samples. The Neural Network method used is backpropagation.

4.2 Recognition Method Algorithm

Neural Network Method (NN) is chosen for voice-based command recognition method because it can handle bigger databases. For Neural Network to implement pattern recognition is quite common, and most of the method used is backpropagation. Supervised learning that starts by inputting the training data through the network is a form of this method. When the data is put in the network, it will generate propagation output activations and then propagated backwards through the neural network, and generating a delta value for all hidden and output neurons. The weights of the network are then updated by calculating delta values that generate by neural network, which increase the speech and quality of the learning process.

For Backpropagation algorithm, the net input to unit i in layer k+1 is shown in Equation 4.1.

$$n^{k+1}(i) = \sum_{j=1}^{Sk} w^{k+1}(i,j)a^{k}(j) + b^{k+1}(i) \tag{4.1}$$

Where

k = is the number of layer

w = is weight

b = is the input of the network

n = is net input

i = is the row of matrix input

j = is the column of matrix input

The output of unit k is shown in Equation 4.2.

$$a^{k}(j) = f^{k+1}(n^{k+1}(i)) \tag{4.2}$$

Where

a = is the output of the network

f = is the final layer of network

For an M layer network the system equations in matrix form are shown in Equation 4.3 and Equation 4.4.

$$a^{0} = p \tag{4.3}$$

$$a^{k+1} = f^{k+1}\left(W^{k+1}a^{k} + b^{k+1}\right), \ k=0,1,\dots, M\text{-}1. \tag{4.4}$$

Where

M = is network

The task of network is to learn associations between a specified set of input-output pairs $\{(p_1, t_1), (p_2, t_2) \dots (p_Q, t_Q)\}$.

The performance index for the network will be shown using Equation 4.5.

$$V = \frac{1}{2} \sum_{q=1}^{Q} \left(t_q - a_q^M\right)^T \left(t_q - a_q^M\right) = \frac{1}{2} \sum_{q=1}^{Q} e_q^T e_q \tag{4.5}$$

Where

V = is the performance index

t = is the target output

a_q^M = is the output of the network, qth is the input, p_q is presented.

For standard backpropagation algorithm, an approximate steepest descent rule will be used. The performance index, V is approximated by Equation 4.6.

$$V = \frac{1}{2} e_q^T e_q \tag{4.6}$$

Where

e_q^T = is the error for the qth input.

Where the total sum of squares is replaced by the squared errors for a single input/output pair. The approximate steepest (gradient) descent algorithm is shown in Equation 4.7 and Equation 4.8.

$$\Delta w^k(i,j) = -\propto \frac{\partial V}{\partial \Delta w^k(i,j)} \tag{4.7}$$

$$\Delta b^k(i) = -\propto \frac{\partial V}{\partial \Delta b^k(i)} \tag{4.8}$$

Where

$\propto$= learning rate

Define Equation 4.9 as the sensitivity of the performance index to changes in the net input of unit $\propto$ in layer.

$$\delta^k(i) = -\propto \frac{\partial V}{\partial n^k(i)} \tag{4.9}$$

Where

δ^k= is the sensitivity of network

Now it can be shown by using Equation 4.1, Equation 4.6 and Equation 4.9, that is shown in Equation 4.10 and Equation 4.11.

$$\frac{\partial V}{\partial \Delta w^k(i,j)} = \frac{\partial V}{\partial \Delta n^k(i)} \frac{\partial V}{\partial \Delta w^k(i,j)} = \delta^k(i)a^{k-1}(j) \tag{4.10}$$

$$\frac{\partial V}{\partial b^k(i)} = \frac{\partial V}{\partial \Delta n^k(i)} \frac{\partial n^k(i)}{\partial b^k(i)} = \delta^k(i) \tag{4.11}$$

It can also shown that the sensitivities satisfy the following recurrence relation in Equation 4.12

$$\delta^k = F^M(n^M)W^{k+1^T}\delta^{k+1} \tag{4.12}$$

Where Equation 4.13 and Equation 4.14

$$\dot{F}^k(\underline{n}^k) = \begin{bmatrix} \dot{f}^k(n^k(1)) & 0 & \cdots & 0 \\ 0 & \dot{f}^k(n^k(2)) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \dot{f}^k(n^k(Sk)) \end{bmatrix} \tag{4.13}$$

$$f^k(n) = \frac{df^k(n)}{dn} \tag{4.14}$$

This recurrence relation is initialized at the final layer shown in Equation 4.15.

$$\delta^M = -F^M(n^M)(t_q - a_q) \tag{4.15}$$

The overall learning algorithm now proceeds as follows; first, propagate the input forward using Equation 4.3 and Equation 4.4; next, propagate the sensitivities back using Equation 4.15 and Equation 4.12; and lastly, update the weights and offset using Equation 4.7, Equation 4.8, Equation 4.10 and Equation 4.11 (Murphy, 2014).

In order to perform the required computation Neural Network is called a threshold logic unit. All the different input from connected neuron is sum together and to determine the correct output activation function is used. The command will be recognized if it threshold is greater than 0.5. The threshold will be round into a logic number if threshold $>$ 0.5 the output value is '1' and if threshold $\leq$ 0.5 the output value is ‘0’. The value ‘1’ indicates that the command is recognise and ‘0’ indicates that the command is not recognise.

4.3 Recognition Using Neural Network

For Neural Network target value, the data is classify into four group or class because it has four command ‘Buka Kipas’, ‘Tutup Kipas’, ‘Buka Lampu’, and ‘Tutup Lampu’. Each of command is a vector having four elements for target value. The target value is set using logic number for example when class 1 is set as ‘1’ the other class will be set as ‘0’and if class 2 is set as ‘1’ the other class will be set as ‘0’. It will be continue until last class. Table 4.1, show Malay command target value.

Table 4.1 Malay Command Target Values

Malay Command	Target Value
Buka Kipas	1000
Tutup Kipas	0100
Buka Lampu	0010
Tutup Lampu	0001

The input used for Neural Network toolbox is the dataset from references model of Malay command and the output target is set as Table 4.1. The data of each sample is 19 x 29 features, the feature will be change into an array and arrange in column using MATLAB. Each Malay command had 30 samples, after the sample is arranged a 551 x 120 dataset is produce. This dataset is the input of Neural Network toolbox. In Neural Network toolbox, the input will be represent by ‘x’ and the output target will be represent by ‘t’. Figure 4.1 shows the input and target insert into Neural Network toolbox.

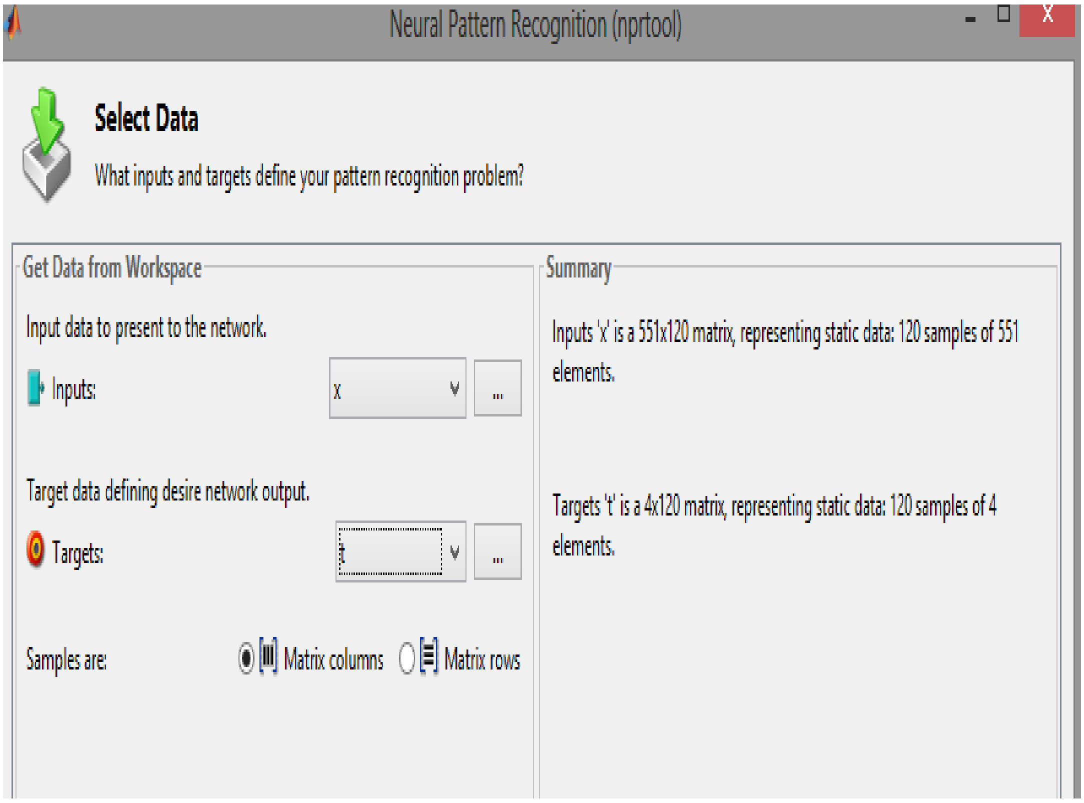

Figure 4.1 The Input and Target insert in Neural Network toolbox

After that, the 'x' is applied to the network and an output is calculated. Then the output will be compare to 't' and assigns values to different classes. As the training algorithm continue, the weight of the network are adjusted in order to minimize the errors in recognition. So the further training iteration continue, the better chance neural network will be assign to a higher value to correct class of the specific input.(Murphy, 2014)

Before build the Neural Network the input data will be divide for training, validation and testing. The division used are 75% samples for training, 10% samples for validation and 15% samples for testing. Training data are represented to the network during training and the network will be adjust according to its error. It is important to compute a model that not only finds a good weight configuration for the training set but also can predict new data with good error. Validation data is a set of data for the function that need to learn, which not directly use to train network. This data is used to minimize overfitting. The weights of the network with this dataset is not adjust but just verify is there any increase in accuracy over the training data set actually yields an increase in accuracy over a data set that not been shown to the network and for testing data, it is a set of examples used only to assed the performance of classifier. This set of data will be not use in training process (Hazry, 2015). Figure 4.2 show selected train, validation and test data.

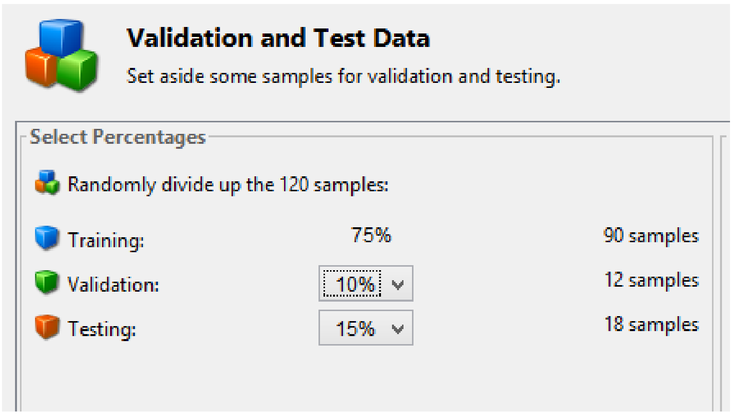

Figure 4.2 Selected train, validation and test data

After choose how many samples need to be train and test, choose the number of hidden layer neural network. The number of hidden layer should be between the size of input layer and the size of output layer. Eighteen of hidden layer neural network is test to see the highest accuracy start from 10 to 90 layers in multiples of five. Then, 90 hidden layers is choose because it has the highest accuracy compare to others. The accuracy of neural network decrease after 90 hidden layer because the classification error is start to increase. Then the Neural Network is train to get the confusion matrix that shown the percentage of the true and predicted classes match. . Table 4.2 show the accuracy of hidden neuron. Figure 4.3 show the structural of neural network.

Table 4.2 Selected of Hidden Neuron

Hidden Neuron	Accuracy (%)
10	87.5
15	93.3
20	94.2
25	95.8
30	92.5
35	90.8
40	91.7
45	94.2
50	91.7
55	96.7
60	90.0
65	94.2
70	94.2
75	90.0
80	91.7
85	94.2
90	98.3

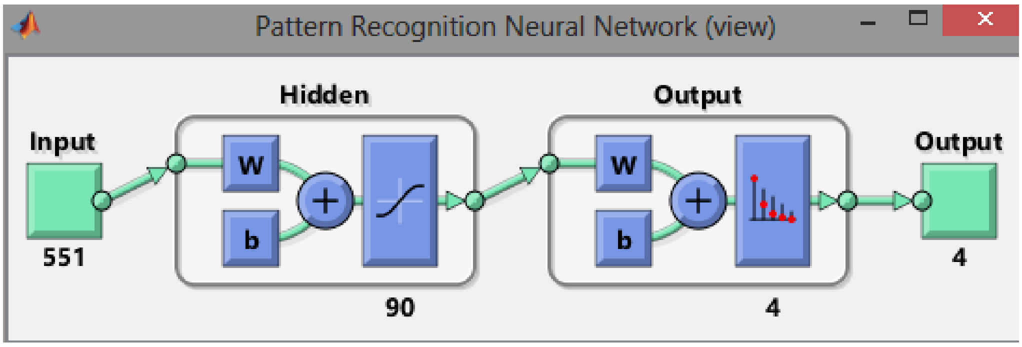

Figure 4.3 Structure of Neural Network

In the confusion matrix, the first four diagonal green cells show the number and the percentage of correct classifications by the trained network. In the confusion matrix 29 commands are correctly classify as 'Buka Kipas'. This corresponds to 24.2% of all 120 command. Same as other classification 30 command correctly classified as 'Tutup Kipas' correspond to 25%, 30 command correctly classified as 'Buka Lampu' correspond to 25% and 29 command correctly classified as 'Tutup Lampu' correspond to 24.2%.

1 of 'Buka Kipas' command is incorrectly classified as 'Buka Lampu' correspond to 0.8% of all 120 command. Same as 1 of the 'Tutup Lampu' command is incorrectly classified as 'Buka Lampu' correspond to 0.8% of all 120 command.

For horizontal cells, 100% are correct out of 29 'Buka Kipas' predictions, 100% are correct out of 30 'Tutup Kipas' predictions, 93.8% are correct out of 32 'Buka Lampu' predictions and 6.3% are wrong and lastly, 100% are correct out of 29 'Tutup Lampu' predictions. For vertical cells, 96.7% are correct out of 30 'Buka Kipas' predictions and 3.3% are wrong, 100% are correct out of 30 'Tutup Kipas' predictions, 100% are correct out of 30 'Buka Lampu' predictions, 96.7% are correct out of 30 'Tutup Lampu' predictions and 3.3% are wrong.

Overall, 98.3% of the prediction are correct and 1.7% are wrong classification. Figure 4.4 show the confusion matrix plot by Neural Network toolbox.

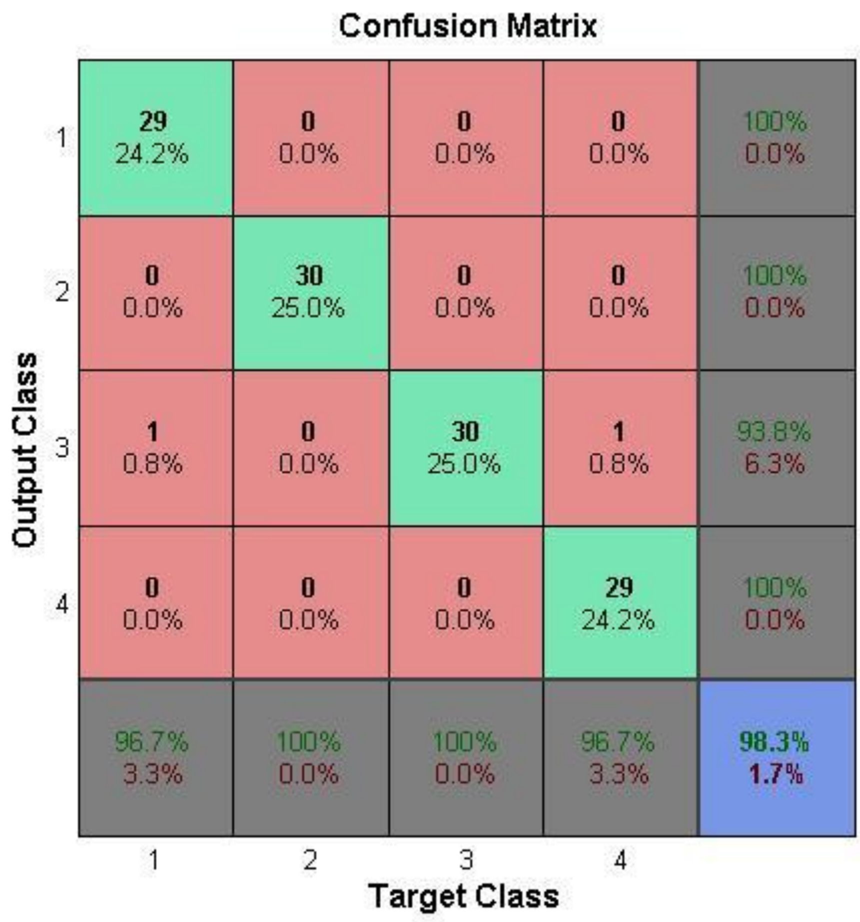

Figure 4.4 Confusion Matrix Plot by Neural Network toolbox

4.4 Performance of Voice-Based Malay Command Recognition

After reference samples had been train with neural network, 720 samples will be test with the reference sample and the system will tell if the command is recognized or not. The data extract from the test sample is compare with the training dataset. The result obtain from simulating training samples in neural network and testing samples is round into '1' or '0' and the command result is execute based on the position of '1' in the target value. Figure 4.5 show the recognition result in MATLAB command window.

```
A =

M1BKNOR (1!).wav

Warning: WAVREAD will be removed in a future release. Use AUDIOREAD instead.
> In wavread at 62
  In mfcc2 at 7
the command given is buka kipas
```

```
A =

M1TKNOR (1!).wav

Warning: WAVREAD will be removed in a future release. Use AUDIOREAD instead.
> In wavread at 62
  In mfcc2 at 7
the command given is tutup kipas
```

```
A =

M1BLNOR (1!).wav

Warning: WAVREAD will be removed in a future release. Use AUDIOREAD instead.
> In wavread at 62
  In mfcc2 at 7
the command given is buka lampu
```

```
A =

M1TLNOR (1!).wav

Warning: WAVREAD will be removed in a future release. Use AUDIOREAD instead.
> In wavread at 62
  In mfcc2 at 7
the command given is tutup lampu
```

Figure 4.5 Recognition result in MATLAB command window

Table 4.3 and table 4.4 show the recognition result of 'Buka Kipas'. The recognition result will be represent in a threshold logic number the command is recognized if the logic number is '1' and the command is not recognized if the logic number is '0'.

Table 4.3 Recognition result for Command Signal 'Buka Kipas' Normal Environment

Speaker	Recognition Result		
	Sample 1	Sample 2	Sample 3
Male 1	1	1	1
Male 2	1	1	1
Male 3	1	1	1
Male 4	1	1	1
Male 5	1	1	0
Male 6	1	1	1
Male 7	1	1	1
Male 8	1	1	1
Male 9	1	1	1
Male 10	1	1	0
Male 11	1	1	1
Male 12	1	0	0
Male 13	1	1	1
Male 14	1	1	1
Male 15	1	1	1
Female 1	1	1	1
Female 2	1	1	1
Female 3	1	1	1
Female 4	1	1	0
Female 5	1	1	1
Female 6	1	0	0
Female 7	1	1	1
Female 8	1	1	1
Female 9	1	0	1
Female 10	1	1	1
Female 11	1	1	1
Female 12	1	0	1
Female 13	1	1	1
Female 14	1	1	1
Female 15	1	0	0

1 = Recognized Command 0 = Not Recognized Command

Table 4.4 Recognition result for Command Signal 'Buka Kipas' Noisy Environment

Speaker	Recognition Result		
	Sample 1	Sample 2	Sample 3
Male 1	1	1	1
Male 2	1	1	0
Male 3	1	1	1
Male 4	1	0	1
Male 5	0	1	0
Male 6	1	0	1
Male 7	1	1	1
Male 8	1	1	0
Male 9	0	0	0
Male 10	0	0	0
Male 11	1	1	0
Male 12	0	0	0
Male 13	0	0	0
Male 14	0	1	0
Male 15	1	1	1
Female 1	1	0	1
Female 2	0	1	1
Female 3	1	1	1
Female 4	0	0	0
Female 5	1	0	1
Female 6	0	0	1
Female 7	1	1	1
Female 8	1	0	0
Female 9	1	1	1
Female 10	1	1	1
Female 11	0	0	1
Female 12	1	1	0
Female 13	1	1	1
Female 14	1	1	1
Female 15	0	1	0

1 = Recognized Command 0 = Not Recognized Command

Based on the result from table 4.3 and table 4.4 the recognition rate is calculate using equation 4.1.

$$\text{Recognition rate} = \frac{N_{correct}}{N_{total}} \times 100\% \qquad \text{Equation 4.1}$$

$N_{correct}$ = the number of correct recognition of testing command samples per command

N_{total} = the total number of testing command samples per command

In normal environment, for 'Buka Kipas' command the correct number of recognise command is 79 samples divide it with 90 samples. The result will be time with 100% so the recognition rate is 88%. For 'Tutup Kipas' command, the correct number of recognise command is 79 samples divide it with 90 samples. The result will be time with 100% so the recognition rate is 88%. The third command is 'Buka Lampu' command, the correct number of recognise command is 79 samples so divide it with 90 samples. The result will be time with 100% so the recognition rate is 88%. Lastly, for 'Tutup Lampu' command the correct number of recognise command is 73 samples so divide it with 90 samples. The result will be time with 100% so the recognition rate is 81%.

In noisy environment, for 'Buka Kipas' command the correct number of recognise command is 54 samples so divide it with 90 samples. The result will be time with 100% so the recognition rate is 60%. For 'Tutup Kipas' command, the correct number of recognise command is 46 samples so divide it with 90 samples. The result will be time with 100% so the recognition rate is 51%. For 'Buka Lampu' command, the correct number of recognise command is 39 samples so divide it with 90 samples. The result will be time with 100% so the recognition rate is 43%. Lastly, for 'Tutup Lampu' command the correct number of recognise command is 18 samples so divide it with 90 samples. The result will be time with 100% so the recognition rate is 20%. Figure 4.6 show the calculated recognition rate in a graph.

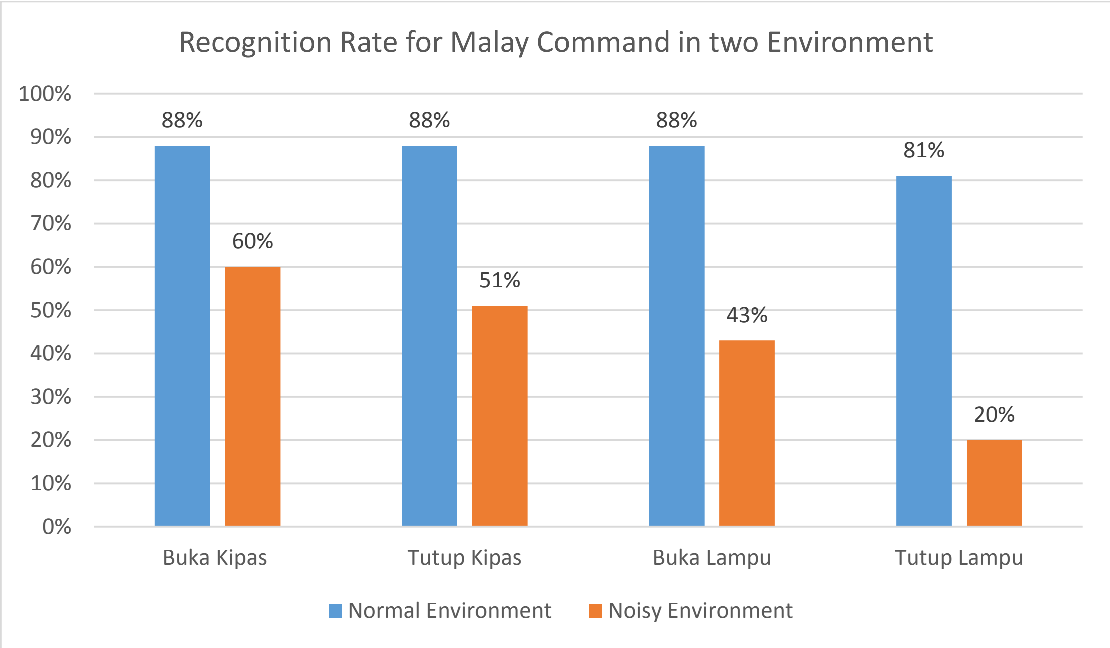

Figure 4.6 Recognition Rate for Malay Command in Normal Environment and Noisy Environment.

The performance of this research is shows in normal and noisy environment. It show that the performance for normal environment is 88% for 'Buka Kipas', 'Tutup Kipas' and 'Buka Lampu' but 81% for 'Tutup lampu'. The performance of normal environment cannot achieve above 90% because some command signal show same pattern for different command. Maybe the command choose is not suitable for this system.

For noisy environment, the performance for 'Buka Kipas' is 60%, 'Tutup Kipas' is 51%, 'Buka Lampu' is 43% and 'Tutup lampu' is 20%. The command cannot achieve 90% because the noise is disturbing the command signal. Therefore, before used this application any background noise should be close or volume should be slow.

To improve, the command choose should be in discontinuous speech instead of continuous speech because the command involve separating sentences. Discontinuous speech will make the command into individual words or phonemes.

Besides that, other also factor for example the health condition of the speaker, speaker speaking rate and speaker slang especially Sabah, Kelantan and Kedah may affect the command signal. Slang for difference races for example Chinese and Indian also can be the factor of signal change.

4.5 Summary

This chapter discuss the results and performance of each command given to electrical appliance using Neural Network. This chapter also explain how to use Neural Network toolbox and how it works. The performance of voice-based Malay command recognition is show in a graph. The graph show that this system perform better in normal environment compare to noisy environment.

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 Conclusion

Voice-based Malay command recognition for electrical appliances has very limited standard Malay for Malay command compare to English. Based on the first objective, some electrical appliance in the living room had been choose to create Malay command database for electrical appliances in the living room. Light and fan had been choose to build the Malay command.

The second objective of this research is to extract the Malay command data by using Mel Frequency Cepstral Coefficient (MFCC) is also achieve. Before extract using MFCC, the command signal is process in pre-processing stage to filter out any unwanted noise. From the data extract, 120 samples is used for training data and then the other data will become testing data.

The third objective is to analyse the Malay command signal using neural network to see the performance of the command in normal environment and noisy environment. The performance will be present in recognition rate.

In conclusion, the important step in voice-based Malay command recognition are speech pre-processing, feature extraction and recognize command. These three method affect the accuracy of the command recognition performance.

5.2 Future Work

For future work, other type of recognition method can be used instead of Neural Network to improve the performance of voice-based Malay command recognition method. Besides that, the distance between speaker and the system should be test instead of fixed the distance so it can tell which distance has better recognition result for the Malay command given.

Lastly, using this recognition method the real-time voice-based Malay command recognition in a living room can be develop by testing if the Malay command given can open and close the electrical appliances in the living room in normal and noisy environment. This system can be develop in a software and hardware. For software, an android can be built as tools to give Malay command so the user can open and close the electrical appliance every in the house. For hardware, a device that resemble electrical appliance in the living room should be develop to see the command performance.

5.3 Budget

Since this project is involved with software development, all the application already provided by Universiti Malaysia Sabah. The main software that I use is MATLAB, this software can be install in Faculty of Engineering laboratory. The other software I use is audacity, this software can be downloaded freely from the webpages.

REFERENCE

Abhay Kumar. 2015. "Voice based control command signal generation for intelligent system." IEEE UP Section Conference on Electrical Computer and Electronics (UPCON)

Alexander Murphy. 11 April 2014. "Implementing Speech Recognition with Artificial Neural Networks."

Aman Ankit, Sonu Kumar Mishra, Rinaz Shaikh, Chandraketu Kumar Gupta, Prakhar Mathur and Soudamini Pawar. 2016. "A Survey Paper on Acoustic Speech Recognition Techniques." ISSN (Online): 2347 - 2812, Volume-4, Issue -7

Fadhilah binti Rosdi. May 2008. "Isolated Malay Speech recognition using hidden markov model."

G. Hemakumar and P. Punitha. 2013. "Speech Recognition Technology: A Survey on Indian Languages."

H. F. Ong and A. M. Ahmad. June 2011. "Malay language speech recognition with hybrid hidden markov model and artificial neural network (HMM/ANN)." International Journal of Information and Education Technology, Vol. 1, No. 2

Jianliang Meng,Junwei Zhang, and Haoquan Zhao. 2012. "Overview of the Speech Recognition Technology."

Jun Tao and Xiaoxiao Jiang. 2011. "A Domestic Speech Recognition Based On Hidden Markov Model."

Malay Culture. 2007. "Malay language in Malaysia" retrieve from www.melayuonline.com/eng/culture/dig/1625.

Md. Ismail Rahman, Bakhteyar Uddin Chowdhury, Raqibul Hossen, Kazi Tanvir Ahmed. 21-23 May 2015. "Multichannel signal detection and selection with android based voice controlled system using simulink." 2nd Int'l Conf. on Electrical Engineering and Infonnation & Communication Technology (ICEEICT) 2015 Jahangirnagar University, Dhaka-1342, Bangladesh,

Mohd Hazry bin Haidir. 2015. "Human Posture Classification"

Mr. Anand Mantri, Mr. Mukesh Tiwari, and Mr. Jaikaran Singh. February 2014. "Performance Evaluation of Human Voice Recognition System based on MFCC feature and HMM classifier." Int. Journal of Engineering Research and Applications www.ijera.com Vol. 4, Issue 2 (Version 1), pp.715-719

Mr.Yoghesh Dawande and Dr. Mukta Dhopeshwarkar. Jan 2015. "Analysis Of Different Feature Extraction Techniques For Speaker Recognition System: a Review." International Journal of Advanced Technology & Engineering Research (IJATER) www.ijater.com, Volume 5, Issue 1

Nihat Ozturk and Ulvi Unozkan. 2010. "Microprocessor based voice recognition system realization."

Norhafizah bt Aripin and M. B. Othman. 2014."voice control of home appliances using android." 39th Annual Northeast Bioengineering Conference

Om Prakash Prabhakar and Navneet Kumar Sahu. May 2013. "A Survey On: Voice Command Recognition Technique."

Parichart Leechor, Chomtip Pornpanomchai, Phichate Sukklay. 2010. "Operation of a radio-controlled car by voice commands."

Pialy Barua, Kanij Ahmad, Ainul Anam Shahjamal Khan, Muhammad Sanaullah. 2014. "Neural Network Based Recognition of Speech Using MFCC Features." 3rd INTERNATIONAL CONFERENCE ON INFORMATICS, ELECTRONICS & VISION 2014

Poonam Sharma and Anjali Garg. May 2016. "Feature Extraction and Recognition of Hindi Spoken Words using Neural Networks." International Journal of Computer Applications (0975 – 8887) Volume 142 – No.7

Rawan T. Khalil, Ala' F. Khalifeh, Khalid A.Darabkh. 2012. "Mobile-free driving with Android phone: system design and performance evaluation."

Samah A.F. Manssor, Abdalla A. Osman and Sally D. Awadalkareem. 2015. "Controlling home devices for handicapped people via command techniques."

Sanjay Bhardwaj, Sunil Pathania and Rajesh Akela. 2015."Speech Recognition using Hidden Markov Model and Viterbi Algorithm."

Shikha Gupta, Mr. Amit Pathak, and Mr.Achal Saraf. August 2014. "A Study of Speech Recognition System: A Literature Review." International Journal of Science, Engineering and Technology Research (IJSETR), Volume 3, Issue 8,

Urmila Shrawankar and Anjali Mahajan. 2013. "Speech User Interface for Information Retrieval."

Wenhao Ou, Wanlin Gao, Zhen Li, Shuliang Zhang, Qing Wang. 2010. "Application of Keywords Speech Recognition in Agricultural Voice Information System."

www.audio.online-convert.com/convert-to-wav

APPENDIX A

Table A.1 Recognition result for Command Signal 'Tutup Kipas' Normal Environment

Speaker	Recognition Result		
	Sample 1	Sample 2	Sample 3
Male 1	1	1	1
Male 2	1	1	1
Male 3	1	1	0
Male 4	1	1	1
Male 5	1	1	1
Male 6	1	1	1
Male 7	1	1	1
Male 8	1	1	1
Male 9	1	1	1
Male 10	1	0	1
Male 11	1	1	1
Male 12	1	1	1
Male 13	1	1	1
Male 14	1	0	1
Male 15	1	1	1
Female 1	1	1	1
Female 2	1	1	1
Female 3	1	1	0
Female 4	1	1	1
Female 5	1	1	0
Female 6	1	1	0
Female 7	1	1	1
Female 8	1	1	0
Female 9	1	1	1
Female 10	1	1	1
Female 11	1	1	0
Female 12	1	1	1
Female 13	1	0	0
Female 14	1	1	1
Female 15	1	0	1

1 = Recognized Command   0 = Not Recognized Command

Table A.2 Recognition result for Command Signal 'Buka Lampu' Normal Environment

Speaker	Recognition Result		
	Sample 1	Sample 2	Sample 3
Male 1	1	1	1
Male 2	1	1	1
Male 3	1	1	1
Male 4	1	1	1
Male 5	1	1	0
Male 6	1	0	1
Male 7	1	1	0
Male 8	1	1	0
Male 9	1	0	1
Male 10	1	1	1
Male 11	1	1	1
Male 12	1	1	1
Male 13	1	1	0
Male 14	1	1	1
Male 15	1	1	1
Female 1	1	1	1
Female 2	1	1	1
Female 3	1	1	1
Female 4	1	1	1
Female 5	1	0	1
Female 6	1	1	1
Female 7	1	0	1
Female 8	1	1	1
Female 9	1	1	0
Female 10	1	1	1
Female 11	1	1	1
Female 12	1	1	1
Female 13	1	0	1
Female 14	1	1	0
Female 15	1	1	1

1 = Recognized Command 0 = Not Recognized Command

Table A.3 Recognition result for Command Signal 'Tutup Lampu' Normal Environment

Speaker	Recognition Result		
	Sample 1	Sample 2	Sample 3
Male 1	1	1	1
Male 2	1	1	1
Male 3	1	1	1
Male 4	1	1	1
Male 5	1	0	1
Male 6	1	0	0
Male 7	1	1	1
Male 8	1	1	1
Male 9	1	1	0
Male 10	1	1	1
Male 11	1	1	1
Male 12	1	1	0
Male 13	1	1	1
Male 14	1	1	1
Male 15	1	1	1
Female 1	1	1	0
Female 2	1	1	1
Female 3	1	1	0
Female 4	1	1	0
Female 5	1	0	0
Female 6	1	0	1
Female 7	1	1	1
Female 8	1	1	1
Female 9	1	1	0
Female 10	1	1	1
Female 11	1	0	0
Female 12	1	0	0
Female 13	1	1	1
Female 14	1	0	1
Female 15	1	1	1

1 = Recognized Command 0 = Not Recognized Command

Table A.4 Recognition result for Command Signal 'Tutup Kipas' Noisy Environment

Speaker	Recognition Result		
	Sample 1	Sample 2	Sample 3
Male 1	1	0	1
Male 2	1	0	1
Male 3	0	1	0
Male 4	0	0	1
Male 5	0	0	0
Male 6	1	1	1
Male 7	0	1	0
Male 8	0	0	1
Male 9	0	0	1
Male 10	0	0	0
Male 11	1	1	1
Male 12	0	1	1
Male 13	1	0	1
Male 14	1	0	0
Male 15	0	0	1
Female 1	1	1	0
Female 2	1	1	1
Female 3	1	1	0
Female 4	1	1	1
Female 5	1	1	0
Female 6	1	1	1
Female 7	0	0	1
Female 8	1	0	1
Female 9	1	0	1
Female 10	1	1	1
Female 11	0	0	0
Female 12	0	0	0
Female 13	1	0	0
Female 14	0	0	0
Female 15	0	0	1

1 = Recognized Command   0 = Not Recognized Command

Table A.5 Recognition result for Command Signal 'Buka Lampu' Noisy Environment

Speaker	Recognition Result		
	Sample 1	Sample 2	Sample 3
Male 1	1	1	1
Male 2	0	1	0
Male 3	0	0	0
Male 4	0	0	1
Male 5	1	1	1
Male 6	0	0	0
Male 7	0	1	1
Male 8	1	1	1
Male 9	0	0	0
Male 10	0	0	0
Male 11	0	0	0
Male 12	0	1	1
Male 13	1	0	0
Male 14	0	0	0
Male 15	0	1	1
Female 1	0	1	1
Female 2	0	1	0
Female 3	1	1	0
Female 4	0	0	0
Female 5	0	0	1
Female 6	1	1	0
Female 7	0	0	0
Female 8	1	1	1
Female 9	0	0	0
Female 10	0	0	0
Female 11	1	0	0
Female 12	1	1	1
Female 13	1	1	1
Female 14	1	1	1
Female 15	0	0	0

1 = Recognized Command   0 = Not Recognized Command

Table A.6 Recognition result for Command Signal 'Tutup Lampu' Noisy Environment

Speaker	Recognition Result		
	Sample 1	Sample 2	Sample 3
Male 1	1	1	1
Male 2	0	0	0
Male 3	0	0	0
Male 4	0	0	1
Male 5	0	0	0
Male 6	0	0	0
Male 7	0	0	0
Male 8	1	0	1
Male 9	1	1	1
Male 10	0	1	0
Male 11	0	0	1
Male 12	0	0	0
Male 13	0	0	0
Male 14	0	0	1
Male 15	0	0	0
Female 1	0	0	0
Female 2	0	0	0
Female 3	0	0	0
Female 4	0	0	0
Female 5	0	0	0
Female 6	1	0	0
Female 7	1	1	0
Female 8	0	0	0
Female 9	0	0	0
Female 10	0	0	0
Female 11	0	0	1
Female 12	0	0	0
Female 13	0	0	0
Female 14	0	0	0
Female 15	0	1	1

1 = Recognized Command 0 = Not Recognized Command

APPENDIX B

ID	FYP2	Start	Finish	Duration
1	Chapter 3 - Methodology	13/2/2017	11/4/2017	42d
2	Record 720 command samples	13/2/2017	2/3/2017	14d
3	Test all possible filter to remove noise from command signal	2/3/2017	10/3/2017	7d
4	Record more command as back up if other command signal cannot be used.	10/3/2017	20/3/2017	7d
5	Extract data from command signal using feature extraction	23/3/2017	11/4/2017	14d
6	Chapter 4 – Result and analysis	11/4/2017	6/6/2017	41d
7	Produce a dataset from data in feature extraction	11/4/2017	28/4/2017	14d
8	Build output target for recognition	28/4/2017	4/5/2017	5d
9	Learn about Neural Network Toolbox	4/5/2017	23/5/2017	14d
10	Chapter 5 – conclusion and recommendation	23/5/2017	25/5/2017	3d

Figure B.1 Gantt Chart FYP2