# FEATURE ENHANCEMENT AND SELECTION METHODS FOR ISOLATED MALAY SPEECH RECOGNITION

by

**CHONG YEN FOOK**
**(1141310649)**

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

**School of Mechatronic Engineering**
**UNIVERSITI MALAYSIA PERLIS**
2017

## ACKNOWLEDGMENT

The successful completion of this thesis work relies on the influence of many people who have generously given their time and energy in specific ways. I take this opportunity to express my gratitude and thanks to each of you who have been a part of this PhD journey.

First and foremost I would like to express my sincere gratitude to the support and supervision of Dr Hariharan Muthusamy and Dr. Vikneswaran Vijean, who brought in this opportunity, for his guidance, encouragement, suggestions and challenges throughout this work from its beginning. Indeed, his inputs and ideas were of immense help in the making of this thesis.

I take this time to thank the Dean, School of Mechatronic Engineering Assoc. Prof. Dr. Abu Hassan bin Abdullah and program chairman Dr. Ruslizam bin Daud for their cooperation and administrative assistance through this course of study.

I would like to express my gratitude and thanks to vice chancellor of UniMAP, Dato' Prof. Dr. Zul Azhar bin Zahid Jamal, for providing me an opportunity in this university and the financial support through SLAI scholarship of UniMAP.

Most importantly, I would like to dedicate my truthful thankfulness to my dearest parents, Mr. Chong Wi Kian and Mdm. Tan Geok Eng, for their unconditional love and patience in raising me up to more than I can be. Special thanks to Ms. Lim Chee Chin who acted as my pillar and supports me all the way in my study, without them this research would never reached to this stage.

Last but not least, I am deeply thankful to different divinities existing in the universe by the principal concept of faith for successfully completing the research.

**TABLE OF CONTENTS**

## LIST OF TABLES

**NO.** **PAGE**

## LIST OF FIGURES

## LIST OF ABBREVIATIONS

| | |
|---|---|
| ELM | Extreme Learning Machines |
| FS | Feature Selection |
| LPC | Linear Prediction Coefficients |
| LPCC | Linear Predictive Cepstral Coefficients |
| WLPCC | Weighted Linear Prediction Cepstral Coefficients |
| MFCC | Mel Frequency Cepstral Coefficients |
| ABC | Artificial Bee Colony |
| DABC | Discrete Artificial Bee Colony |
| PSO | Particle Swarm Optimization |
| BPSO | Binary Particle Swarm Optimization |
| SD | Subject Dependent |
| SI | Subject Independent |

## LIST OF SYMBOLS

| | |
|---|---|
| µ | Emigration |
| V | Velocity of the Particle |
| P | Population |
| X | Food Source |
| | Cost Function |

**Feature Enhancement and Selection Methods for Isolated Malay Speech Recognition**


## ABSTRACT

Automatic speech recognition (ASR) is a technique to translate automatically incoming speech signal into their contextual information. In the pass few decade, various acoustic feature extraction and classification algorithms have been developed for native English speech recognition and different languages spoken around the world using acoustic signals. Research in Automatic Speech Recognition (ASR) by machines had been done for more than five decades. Various research findings have been reported in recent years in speech recognition for many different languages. However, every languages having their own unique words structure. As examples, English words are formed due to the changes of phoneme in the based word itself according to its group of words and Malay words allow addition of affixes to the base word to form new words. In this research, signal processing techniques are applied to the acoustic signals in an effort to recognize the Malay speech. To reduce the misclassification, the recorded speech signals were segmented to remove the unvoiced speech (noise). In this research works, parametric Linear Prediction Coefficients (LPC), Linear Prediction Cepstral Coefficient (LPCC), Weighted Linear Prediction Coefficients (WLPCC), Mel-Frequency Cepstral Coefficients (MFCC) and non-parametric Wavelet Packet Transform based Energy and Entropy (WPT-EE) representations of features were extracted. The features extracted were enhanced to increase the discriminant ability using artificial bee colony based clustering. Then, the enhanced features set were dimensionally reduced by using two feature selection techniques. They are binary particle swarm optimization (BPSO) and discrete artificial bee colony (DABC) feature selection technique. Last, two classifiers as the probabilistic neural network (PNN) and extreme learning machine (ELM) were used to evaluate the performance of extracted and enhanced features from recorded Malay speech signal. The proposed artificial bee colony based feature enhancement (ABC-FE) features show promising average results of 99.61% (Speaker Dependent) and 96.21% (Speaker Independent). Experimental results showed that the average accuracy obtained by using hybrid features of LPC, LPCC, WLPCC, MFCC and WPT-EE for Speaker Dependent and Speaker Independent with ELM classifier were 97.89% (PSO)-98 features and 99.33% (ABC)-67 features for Malay speech recognition.

**Peningkatan Ciri-ciri dan Kaedah Pilihan untuk Pengecaman Pertuturan Melayu Terpinggir**


## ABSTRAK

Pengecaman pertuturan automatik (ASR) adalah satu teknik untuk menterjemahkan isyarat pertuturan secara automatik masuk ke dalam maklumat konteks mereka. Dalam beberapa dekad yang lepas, pelbagai ciri-ciri akustik pengekstrakan dan klasifikasi algoritma telah dibangunkan untuk pengecaman pertuturan bahasa Inggeris ibunda dan bahasa yang berlainan dituturkan seluruh dunia menggunakan isyarat akustik. Penyelidikan dalam pengecaman pertuturan automatik (ASR) oleh mesin telah dilakukan lebih daripada lima dekad. Pelbagai penemuan penyelidikan telah dilaporkan kebelakangan dalam pengecaman pertuturan untuk pelbagai bahasa. Walau bagaimanapun, setiap bahasa mempunyai struktur perkataan unik mereka tersendiri. Sebagai contoh, perkataan English terbentuk disebabkan perubahan fonem dalam perkataan berdasarkan sendiri mengikut kumpulan kata-kata dan perkataan Melayu membolehkan penambahan imbuhan kepada perkataan asas untuk membentuk perkataan baru. Dalam kajian ini, teknik pemprosesan isyarat digunakan untuk isyarat akustik dalam usaha mengecam pertuturan Melayu. Untuk mengurangkan kesilapan klasifikasi itu, isyarat pertuturan yang direkodkan telah dibahagikan untuk mengeluarkan pertuturan tak bersuara (bunyi bising). Dalam kerja-kerja penyelidikan, parametrik Linear Pekali Ramalan (LPC), Linear Ramalan Cepstral Pekali (LPCC), Wajaran Pekali Ramalan Linear (WLPCC), Mel Frekuensi Cepstral Pekali (MFCC) dan bukan parametrik Wavelet Packet Transform Tenaga dan Entropy (WPT-EE) perwakilan ciri-ciri yang telah diekstrak. Ciri-ciri yang diekstrak telah dipertingkatkan keupayaan diskriminasinya dengan menggunakan pengelompokan berdasarkan tiruan koloni lebah. Kemudian, set ciri-ciri yang telah dipertingkatkan akan dikurangkan secara dimensi dengan menggunakan dua teknik pemilihan ciri. Mereka adalah teknik pemilihan ciri zarah binari swarm pengoptimuman (BPSO) dan teknik pemilihan ciri diskret koloni lebah tiruan (DABC). Lalu, dua pengelas iaitu Rangkaian Kebarangkalian Neural (PNN) dan Mesin Pembelajaran yang Melampau (ELM) telah digunakan untuk menilai prestasi pengekstrakan dan peningkatan ciri-ciri yag direkodkan daripada isyarat pertuturan Melayu. Koloni lebah tiruan yang dicadangkan berdasarkan ciri-ciri peningkatan (ABC-FE) ciri-ciri menunjukkan hasil purata ketepatan sebanyak 99.61% (Penceramah Tangungan) dan 96.21% (Penceramah Bebas). Keputusan eksperimen menunjukkan bahawa ketepatan purata diperolehi dengan menggunakan ciri-ciri hibrid LPC, LPCC, WLPCC, MFCC dan WPT-EE untuk penceramah tanggungan dan penceramah bebas dengan pengelas ELM adalah 97.89% (PSO)-98 ciri-ciri dan 99,33% (ABC) -67 ciri-ciri untuk pengecaman pertuturan Melayu.

# CHAPTER 1 : INTRODUCTION

This chapter presents an introduction to the automatic speech recognition, discussion on the existing methods of Malay speech recognition, its drawbacks and also the advantages of using proposed feature enhancement technique and features selection methods. The objectives of the proposed research and organization of the thesis are also described in the successive section.

## 1.1 Research Background

Speech recognition is a methodology to translate spoken language into their contextual information (Al-Haddad, Samad, Hussain, Ishak, & Noor, 2009; Ariff, Alwi, & Salleh, 2005; Ong & Ahmad, 2011). Various acoustic feature extraction and classification algorithms have been developed for native English speech recognition and different languages spoken around the world using acoustic signals. Research in speech recognition by machines had been done for more than five decades. Over the past decades, the development of speech recognition applications give important contributions to this field of research and is becoming more mature in recent years.



Nowadays, computer is a technology that is further than its original design in helping human to solve problem. The mouse and keyboard are usually used as an input and the monitor and printer as an output by computer (Rahman, Mohamed, Mustafa, & Salim, 2014). In fact, speech has the potential to be a better and easier medium for human to interact with computer (Rosdi & Ainon, 2008). Various research findings have been reported in recent years in speech recognition for many different languages (Kassim, Maarof, Zainal, & Wahab, 2016a; Lee, Low, & Mohamed, 2013).

However, every language have their unique words structure. Examples, English words are formed due to the changes of phoneme in the based word itself according to its group of words and Malay words allow addition of affixes to the base word to form new words (Ting, Yunus, Salleh, & Cheah, 2001).



Speech recognition can be approximately divided into two stages: feature extraction and recognition. Feature extraction is important as a step to minimize the dimensionality of the input data, and extract the useful information from the input data. In typical speech feature extraction technique, speech signals were segmented into frames and extracted features from each frame. Through feature extraction algorithms, speech signals are changed into a sequence of feature vectors (Majeed, Husain, Samad, & Hussain, 2012). Then these feature vectors are used in the recognition stage. Information loss during the transition from speech signals to a sequence of feature vectors can be minimized. Feature extraction stage plays a very important role for speech recognition performance. This is the result of the fact that an effective feature vector gives rise to better recognition performance (Avci, 2007).

Mel-Frequency Cepstral Coefficients (MFCC), and Linear Predictive Cepstral Coefficients (LPCC) techniques are famously used at the feature extraction stage in speech recognition because of their ability to estimate the process of human hearing and perceive sounds with different frequencies. Initially, at the preprocessed stage which contains the steps of sampling, quantification, windowing, end-point detection and filtering are applied to the input acoustic signal (Kaur & Singh, 2012; Liu, He, & Palm, 1996; Ooi, Hariharan, Yaacob, & Lim, 2012; Seman, Bakar, & Bakar, 2010c). Thus, both LPCC and MFCC have been used in this research.

In recognition stage several approaches have been reported such as Hidden Markov Models (HMM), Dynamic Time Warping (DTW), Support Vector Machines (SVM), Artificial Neural Networks (ANN). Based on previous works, neural networks have proven to be more efficient in speech recognition (Al-Haddad, Samad, Hussain, Ishak, & Mirvaziri, 2007; Huang, Zhu, & Siew, 2004; Kalamani, Valarmathy, & Anitha, 2015; Muthusamy, Polat, & Yaacob, 2015; H N Ting, Yunus, & Nordin, 2005). Thus, both Probability Neural Network (PNN) and Extreme Learning Machines (ELM) have been used in this research.

In addition, current Malay speech recognition is quite sensitive to environment and style of speech variations. When we step outside the boundaries, a number of techniques for improving ASR robustness have met limited success in severely degraded environments in which ambient noise presence. In this research, various signals processing techniques are used for Malay speech signals in an effort to enhance the Malay speech recognition (Hernando, Nadeu, & Mareo, 1997; Liu, Zhao, Pi, Liang, & Nefian, 2002; Potamianos, Neti, Luettin, & Matthews, 2004; Rosdi & Ainon, 2008; Salam, Mohamad, & Salleh, 2011; Seman, Bakar, & Bakar, 2010a).

## 1.2 Problem Statement

The present state of Speech Recognition is limited and sensitive to environment channels (Potamianos et al., 2004). We know that speech recognition had been established more than five decades, but present researches are mostly focused on their own native languages like English, Japanese, Spanish and German. Here, Malay speech was focused in this research. For Malay speech recognition, the existing MSR were only investigated on parametric features (LPC, LPCC, WLPCC, and MFCC) alone from acoustic signal which are high possibility prone to misclassification and poor recognition due to environment factors (presence of ambient noise) (Rosdi & Ainon, 2008; Seman & Jusoff, 2008). Moreover, the environment is another critical issue for speech recognition as the researcher is required to get the input speech in clean environment. In this research work, the hybrid between time-domain feature (LPC, LPCC, WLPCC and MFCC) and time-frequency-domain feature (wavelet packet transform based energy and entropy features) were proposed. In this research work, the performance will be evaluated using different signal-to-noise (SNR) ratio environment. This is to ensure good precise and high performance of Malay speech recognition that adapts to multi-environments for future development and application purpose.



## 1.3 Research Objectives

The aim of this research is to design and develop an intelligent Malay speech recognition using speech signals. The objectives of this research are as follows.

i) To develop Isolated Malay Word Database (Digits and Words)

ii) To develop pre-processing and feature extraction methods for extracting salient features from the recorded spoken utterances.

iii) To enhance the discrimination ability of the features using the proposed Meta-heuristics based Clustering and feature selection methods.

iv) To evaluate the proposed methods in recognizing Isolated Malay Words.

## 1.4 Research Scope

Development of robust Malay speech recognition is still a challenging issue because of noise and complexity of Malay speech structure. Therefore, this research work is an effort to extract the salient features from the recorded spoken utterances to represent the Malay speech. A dataset that consists 10 Malay digits ('KOSONG' to 'SEMBILAN') was created to evaluate the effectiveness of parametric and non-parametric feature representation which included LPC, LPCC, WLPCC, MFCC and Wavelet Packet Transform based energy and entropy features. After that, a dataset that consists of 50 Malay words which included *vowels*, *diphthongs*, *primary consonants* and *secondary consonants*. The database consists of 24 subjects (6 males and 18 females) with age range between 20 and 28 years old and it will be used in this research. The effectiveness of Malay speech recognition with proposed methods will be investigated in clean and noisy level in range between -10dB and 30dB.

Moreover, artificial bee colony clustering technique was used to enhance the class separation between classes. After that, both discrete artificial bee colony (DABC) and binary particle swam optimization (BPSO) algorithms will be used as feature selection in this research work to optimize the useful features from hybrid features set. Lastly, both PNN and ELM classifiers will be used to evaluate the proposed methods in recognizing isolated Malay Words.

## 1.5 Thesis Organization

In this research work, two non-linear classifiers are used (PNN classifier and ELM classifier). Their performance on the recognition of Malay Speech Recognition is reported. The proposed research carried out in order to address the four objectives is presented in five chapters of this thesis.

Chapter 1 (current chapter) explains the introduction of speech recognition. This chapter discusses the limitation of existing methods of Malay speech recognition. This chapter also deals with the objectives of the proposed research, problem statement, research scope, and thesis organization.

Chapter 2 describes the introduction to Malay speech recognition, speech production mechanism and linguistic information about Standard Malay (SM). This chapter gives the review of significant works by previous researchers on Malay speech recognition.

Chapter 3 describes the material and method which includes dataset used, Malay speech signals collection procedure, speech signals processing algorithms and classification algorithms were used throughout this research.

Chapter 4 presents the efficacy of the proposed feature enhancement method and selection methods for the Malay speech recognition under normal and various noisy conditions.

Chapter 5 summarizes the overall research findings, contributions and presents the future planning and the direction of the research work.

# CHAPTER 2 : LITERATURE REVIEW

## 2.1 Automatic Speech Recognition

Speech recognition is a technology allows a machine to converts a speaker's spoken utterance into a text or used for other applications (Anusuya & Katti, 2009). Automatic speech recognition (ASR) system has been a major topic of interest in speech research for more than five decades (Rabiner & Juang, 1993). Unfortunately, speech recognition is still far from a solved problem (Al-Haddad et al., 2009).

Nowadays, speech technology is widely used by people in their daily life. People are using speech input to send email, open an internet browser and search information in internet and perform some tasks in their computer. Moreover, the speech inputs are used by disabled people to operate certain tasks in their daily work (Salam, Mohamad, & Salleh, 2011). For example, it is now possible to type a document with the help of speech input into their computer machine. Also, human-machine interface such as telephone call, flight ticket booking, restaurant reservation, and food ordering system have also been successfully implemented with the help of speech technology today.



Overall, this chapter presents a review of speech recognition assignment, speech recognition methodologies, current speech recognition and Malay speech recognition as well as different type of approaches applied in Malay speech recognition.

## 2.2 Human Speech Production Mechanism

Speech is essential for people to communicate with each other. Speech translates different sorts of information which is significant to express point of view and information conveyed. Most essential components of the human speech production framework are the lungs, trachea, larynx or its vocal cords, nasal cavity, soft palate or velum, hard palate, tongue, teeth and lips (Karpov, 2003).

### 2.2.1 Anatomy of Speech Production

The speech production procedure is to release the air from the lungs into the vocal tract (Rabiner, 1978). The air flow created by the lungs propagates over the larynx, the pharynx and the upper vocal tract as shown in Figure 2.1.

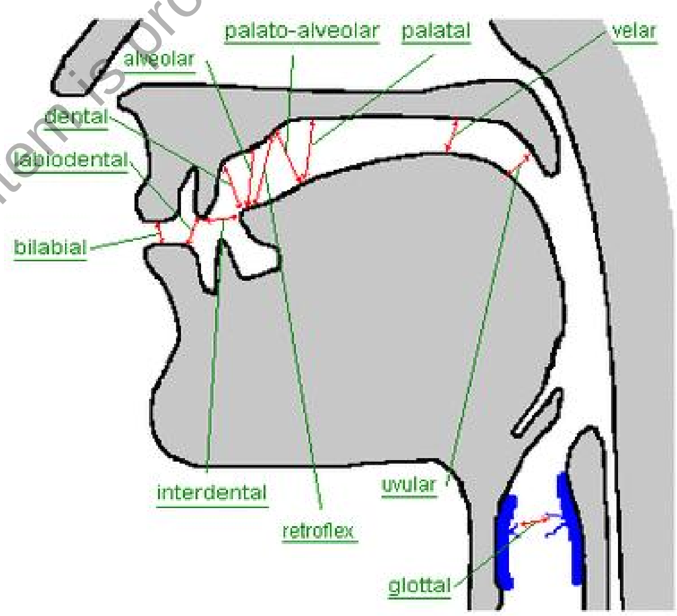


Figure 2.1: Human speech production mechanism (Prator & Robinett, 1985)