MALAY CONTINUOUS SPEECH RECOGNITION USING CONTINUOUS DENSITY HIDDEN MARKOV MODEL

TING CHEE MING

A thesis submitted
in fulfilment of the requirements for the award of the
Degree of Master of Engineering (Electrical)

Faculty of Electrical Engineering
Universiti Teknologi Malaysia

MAY 2007

Dedicated to *Buddha, Dharmma, Sangha*, and my
Beloved Dad, Mum, Sisters & friends.

ACKNOWLEDGEMENT

First of all, I wish to thank heartedly my supervisor Prof. Sheikh Hussain. He is the one who introduce me to the high-tech field of automatic speech recognition. It is his endless guidance whether technical knowledge or ways of conducting research, that build up my foundation on speech recognition field. His moral and financial support survives me during my master program. His attitude and enthusiasm in doing research and his consistent vision to make high-tech research grounded in reality, applicable, and even commercialized, always inspire me.

Secondly, I am grateful to all my research colleagues at Center of Bio-medical Engineering (CBE), Amar, Tan, Kamarul, and others. Thanks you for your generosity to share the resource and knowledge with me. Special thanks to CBE for providing me resource and conducive environment to make my research a success.

Finally, I wish to thank my family and friends for their support. I am especially grateful for my parents for all their sacrifices in upbringing me. The encouragement and support for me in pursuing my research career are appreciated sincerely. I would like to thank my friend Sui Feng, she always by my side, comfort me when I am discouraged, take care of me when I am busy, and share my happiness when I delighted.

ABSTRACT

This thesis describes the investigation of the use of Continuous Density Hidden Markov Model (CDHMM) for Malay Automatic Speech Recognition (ASR). The goal of this thesis is to solve the constraints of current Malay ASR that are: speaker-dependent, small vocabulary and isolated words, and provides a basis in developing speaker-independent (SI) Malay large vocabulary continuous speech recognition (LVCSR). Hidden Markov Model (HMM) based statistical modeling is used in Malay speech recognition. HMM is a robust and powerful technique capable of modeling of speech signals. With their efficient training algorithm (Baum-Welch and Viterbi/Segmental K-mean) and recognition algorithm (Viterbi), as well as it's modeling flexibility in model topology, observation probability distribution, representation of speech unit and other knowledge sources, HMM has been successfully applied in solving various tasks in this thesis. CDHMM which model the continuous acoustic space eliminates quantization error imposed by discrete HMM. CDHMM performs better than discrete HMM in Malay speech recognition. CDHMM with mixture densities which is capable to model inter-speaker variability performs well in multi speaker task (99% in isolated words task). The result expects its well performance in SI task in the future. A connected words ASR is developed and evaluated on Malay connected digit task and has achieved reasonably good accuracy with limited training data. Segmental K-mean training procedure is proven to perform better than the manual segmentation. The sub-word unit modeling is attempted in Malay phonetic classification and segmentation on medium vocabulary Malay continuous speech database. Experiments are conducted to investigate different feature set and mixture components. The knowledge of continuous ASR architecture and sub-word unit modeling gained in this work has provided basis for Malay LVCSR. For conclusion, the basic idea of HMM implemented in other language domain can be successfully applied in the Malay language domain as well.

ABSTRAK

Tesis ini mengkaji *Continuous Density Hidden Markov Mode*l (CDHMM) untuk Sistem Pengecaman Suara (ASR) Melayu. Kajian ini bertujuan untuk mengatasi kelemahan ASR Melayu terkini, dari segi penutur-bersandar, vokabulari kecil dan perkataan berasingan, dan menyediakan asas untuk membangunkan Pengecaman Suara Berterusan Vokabulari Besar (LVCSR) Melayu yang penutur-bebas (SI). Satu model berstatistik iaitu *Hidden Markov Model* (HMM), digunakan dalam ASR Melayu. HMM ialah teknik yang berkesan dalam pemodelan suara, kerana ia mempunyai algoritma latihan (*Baum-Welch and Viterbi/Segmental K-mean*) dan algoritma pengecaman yang berkesan, serta pemodelan yang fleksible pada topologi, serakan kebarangkalian keluaran, perwakilan unit suara dan pengetahuan bagi punca lain. HMM yang diaplikasikan ini, telah berjaya mengatasi pelbagai kerja dalam tesis ini. CDHMM yang memodelkan ruang akustik berterusan dapat menghapuskan masalah kuantasasi yang disebabkan oleh HMM diskrit. CDHMM adalah lebih tepat daripada HMM diskrit dalam ASR Melayu. CDHMM yang mengunakan densiti bergabung, memodelkan variasi antara-penutur, berkesan dalam kerja pelbagai penutur (99% dalam kerja perkataan terasing). Hasil kajian ini menjangka akan berkesan dalam kerja SI. ASR perkataan berhubung yang dibangunkan dan dinilai dalam kerja digit Melayu berhubung, mencapai ketepatan yang memuaskan dalam data latihan yang terhad. Prosedur latihan *Segmental K-mean* disahkan lebih tepat daripada sekmentasi insani. Permodelan unit sub-perkataan dikaji dalam pengelasan dan sekmentasi phonem Melayu untuk pengkalan data suara berterusan Melayu yang bervokabulari serderhana. Experimen dijalankan untuk mengkaji pelbagai ciri dan komponen bergabung. Pengetahuan dalam struktur ASR berterusan and permodelan unit sub-perkataan menjadi asas untuk membina LVCSR Melayu. Sebagai kesimpulan, konsep dasar HMM yang digunakan dalam bahasa lain juga boleh digunakan secara berkesan dalam domain bahasa Melayu.

TABLE OF CONTENTS

CHAPTER	TITLE	PAGE
	TITLE PAGE	i
	ADMISSION PAGE	ii
	DEDICATION	iii
	ACKNOWLEDGEMENT	iv
	ABSTRACT	v
	ABSTRAK	vi
	TABLE OF CONTENTS	vii
	LIST OF TABLES	xiii
	LIST OF FIGURES	xv
	LIST OF SYMBOLS	xviii
	LIST OF APPENDIXES	xxi
1	**INTRODUCTION**	1
	1.1 Introduction	1
	1.2 Problem Statement	3
	1.3 Objectives of the Research	4
	1.4 Scope of Research	5
	1.5 Outline of Thesis	7
	1.6 Contribution of the Thesis	9
2	**LITERATURE REVIEW**	**11**
	2.1 Introduction	11

2.2 Classification of Speech Recognition Task 12
2.2.1 Speaker Dependent versus Speaker Independent 12
2.2.2 Isolated words and Continuous Speech 14
2.2.3 Small Vocabulary versus Large Vocabulary 16
2.3 Automatic Speech Recognition System 17
2.3.1 Discussion of Profound ASR System 17
2.3.2 Discussion of Malay ASR System 19
2.3.3 Speech Recognition Approaches 21
2.3.3.1 Acoustic-Phonetic Approach 22
2.3.3.2 Statistical Pattern Recognition Approach 24
2.3.3.3 Artificial Intelligence Approach 27
2.4 Modeling and Classification Techniques in Speech Recognition 29
2.4.1 Dynamic Time Warping 29
2.4.2 Hidden Markov Model 30
2.4.2.1 Model Topology 31
2.4.2.2 HMM Parameter Re-estimation 32
2.4.2.3 HMM Training Strategies 33
2.4.2.4 HMM Density Architecture 34
2.4.2.5 Speech Units Modeling 36
2.4.2.6 Decoding 38
2.4.3 Neural Network 39
2.4.4 Support Vector Machine 41
2.5 Summary 42

3	**STATISTICAL SPEECH RECOGNITION SYSTEM AND HIDDEN MARKOV MODELING OF SPEECH**		**43**
	3.1	Introduction	43
	3.2	Statistical Speech Recognition	44
		3.2.1 Acoustic Front End Processing	48
		3.2.2 Acoustic Modeling	49
		3.2.3 Language Modeling	52
		3.2.4 Pronunciation Modeling	56
		3.2.5 Decoding	57
	3.3	Feature Extraction – Mel-Frequency Cepstral Coefficient	57
		3.3.1 Mel-Frequency Cepstral Coefficient	57
		3.3.2 MFCC Front End Processing	58
	3.4	Hidden Markov Modeling of Speech	63
		3.4.1 Modeling of Speech	63
		3.4.2 Definition of HMM	64
		3.4.3 HMM Formulation	65
		3.4.4 The Three Problems of HMM	66
		3.4.5 Solutions to the Problems	69
		3.4.5.1 Solution to Problem 1: Forward-Backward Algorithm	70
		3.4.5.2 Solution to Problem 2: Viterbi Algorithm	73
		3.4.5.3 Solution to Problem 3: Baum-Welch Re-estimation Algorithm	76
		3.4.6 Implementation Issues	81
		3.4.6.1 Scaling	81
		3.4.6.2 Multiple Observation Sequences	84

	3.4.7	Strengths of HMM in Modeling Speech	85
		3.4.7.1 Model Topology	86
		3.4.7.2 Observation Probability Distributions	87
		3.4.7.3 HMM Representation of Speech Units	96
		3.4.7.4 HMM Representation of Other Knowledge Sources	99
		3.4.7.5 The Training Algorithms for HMM's	103
4		**EXPERIMENTAL RESULTS AND DISCUSSION ON ISOLATED WORD RECOGNITION**	**111**
	4.1	Introduction	111
	4.2	Isolated Word Recognition System	112
	4.2.1	Training	112
	4.2.2	Recognition	113
	4.3	System Evaluation	114
	4.3.1	Task and Database	114
	4.3.2	Comparison of DHMM and CDHMM	114
		4.3.2.1 Result & Discussion	115
	4.3.3	Comparison with Other Systems	116
	4.4	Summary	118
5		**EXPERIMENTAL RESULTS AND DISCUSSION ON CONNECTED WORDS RECOGNITION**	**119**

	5.1	Introduction	119
	5.2	Connected Words Recognition System	120
		5.2.1 Acoustic Model	120
		5.2.2 Training	120
		5.2.3 Language HMM Construction	123
		5.2.4 Recognition	124
	5.3	System Evaluation	126
		5.3.1 Task and Database	126
		5.3.2 Comparison of Segmental K-mean and Manual Segmentation Algorithm	126
		5.3.2.1 Results for SS Test	127
		5.3.2.2 Results for MS Test	129
		5.3.3 Comparison with Other Systems	130
	5.4	Summary	132
6		**EXPERIMENTS AND DISCUSSION ON PHONETIC CLASSIFICATION AND SEGMENTATION**	**133**
	6.1	Introduction	133
	6.2	CDHMM Phone Modeling	134
	6.3	Malay Phone Set	135
	6.4	Phoneme Classification	136
		6.4.1 HMM classifier	136
	6.5	Phoneme Classification Experiment	137
		6.5.1 Database and Analysis Condition	138
		6.5.2 Effect of Different Feature Set	139
		6.5.3 Effect of Varying Number of Mixture Components	140
		6.5.4 All-phone & Intra-category Classification	141
		6.5.5 Error Analysis	142

	6.6	Phonetic Segmentation	143
		6.6.1 Related Word in HMM Automatic Phonetic Segmentation	144
		6.6.2 HMM based Automatic Phonetic Segmentation	146
	6.7	Phonetic Segmentation Experiment	148
		6.7.1 Database and Analysis Condition	149
		6.7.2 Performance Evaluation Criterion	149
		6.7.3 Effect of Different Feature Set and Varying Number of Gaussian Mixture Components.	150
		6.7.4 Comparison with Other Studies	151
	6.8	Summary	153
7	**SUMMARY AND CONCLUSION**		**154**
	7.1	CDHMM Malay Speech Recognition	154
	7.2	Future Work	157
REFERENCES			**161**
Appendix A-C			176-203

LIST OF TABLES

TABLES NO.	TITLE	PAGE
2.1	Important Malay ASR researches and their freedom from the task constraints.	20
4.1	DHMM and CDHMM recognizers with different training algorithms.	114
4.2	Comparison of DHMM and CDHMM recognizers with different training algorithm.	115
4.3	Comparisons between the isolated word system in this thesis and other similar systems.	117
5.1	Word accuracy for SS test.	128
5.2	Performance comparison of segmental k-mean and manual segmentation algorithm for different language models in MS test.	129
5.3	Comparisons between the connected word system in this thesis and other similar systems.	130
6.1	List of Malay phones according to categories.	135
6.2	Classification accuracy with various combinations of	

	features.	139
6.3	Effect of different number of mixture components on classification accuracy.	140
6.4	Intra-category and all phoneme classification accuracy for different phone category.	141
6.5	Most common errors in all-phoneme classification.	142
6.6	Percentage of boundaries within different tolerances (in ms) with varying number of mixture components using different feature set.	150
6.7	Comparison of HMM based phonetic segmentation studies	152

LIST OF FIGURES

FIGURES NO.	TITLE	PAGE
2.1	Block diagram of the acoustic-phonetic approach for ASR.	22
2.2	Segmentation and labeling for word sequence “seven-six”.	23
2.3	Block diagram of pattern recognition speech recognizer.	25
2.4	Bottom-up approach to knowledge integration	28
2.5	Top-down approach to knowledge integration.	28
2.6	Schematic of basic 3 states left-to-right HMM model.	31
3.1	Schematic overview of statistical speech recognition using Bayes decision rule.	46
3.2	Components of a continuous speech recognition system.	48
3.3	A hierarchically structured HMM.	51
3.4	Acoustic Front End using MFCCs, energy, and differenced coefficients.	58

3.5	A First-order three-state Markov Chain with associated processes.	65
3.6	Illustration of the sequence of operations in computing forward variable.	71
3.7	Illustration of the sequence of operations in computing backward variable.	73
3.8	Illustration of the sequence of operations required for the computation of the joint event that the system is in state S_i, at time t and state S_j, at time $t+1$.	78
3.9	(a) left-to-right HMM and (b) ergodic HMM.	87
3.10	An HMM representing a word.	97
3.11	An HMM representing a phoneme	98
3.12	(a) Language graph for a simple language with a two word vocabulary. (b) Language phonetic graph derived from (a). (c) Language-phonetic-acoustic graph derived from (b). (d)Trellis formed as the cross-product of (c) and a linear graph of observation vectors.	100
3.13	A HMM network for connected word recognition with unigram.	101
3.14	A HMM network for connected word recognition with bi-gram.	102
3.15	A HMM network for continuous digit recognition with unigram.	102

3.16	CDHMM training with BW algorithm.	104
3.17	HMM training with Viterbi algorithm.	108
4.1	Combination of BW and Segmental K-means training algorithm.	113
5.1	The Segmental K-Mean training algorithm for connected word strings.	122
5.2	The HMM based one pass connected words recognition. The recognized words sequence is W1, W4, W3, and W2.	125
6.1	MAP phoneme classifier.	137
6.2	Block diagram of automatic segmentation system.	147
6.3	Viterbi forced alignment in phonetic segmentation. The waveform word “saya” is segmented corresponding to its phone labels.	148

LIST OF SYMBOLS

AI	-	Artificial intelligence
AP	-	Acoustic-phonetic
ASR	-	Automatic Speech Recognition
BW	-	Baum-Welch
CD	-	Context-Dependent
CDHMM	-	Continuous Density Hidden Markov Model
CI	-	Context-Independent
DFT	-	Discrete Fourier Transform
DHMM	-	Discrete Hidden Markov Model
DTW	-	Dynamic Time Warping
E	-	Energy
EM	-	Expectation-Maximization
HMM	-	Hidden Markov Models
KS	-	Knowledge sources
LBG	-	Linde-Buzo-Gray
LM	-	Language Model
LPC	-	Linear Predictive Coding
LPCC	-	Linear Predictive Coding Cepstrum
LVCSR	-	Large vocabulary continuous speech recognition.
LVQ	-	Learning vector quantization

MCE	-	Minimum classification error
MDI	-	Minimum discrimination information
MFCC	-	Mel-Frequency Cepstral Coefficients
MFPC	-	Mel-Frequency Power Cepstrums
ML	-	Maximum Likelihood
MLP	-	Multi-Layer Perceptron
MMI	-	Maximum mutual information
NN	-	Neural Network
PLP	-	Perceptual Linear Prediction
RNN	-	Recurrent neural networks
SCHMM	-	Semi-Continuous Hidden Markov Model
SD	-	speaker-dependent
SI	-	speaker-independent
SOM	-	Relf-organizing maps
SRM	-	Structural risk minimization
SVM	-	Support Vector Machine
TDNN	-	Time delay neural network
VB	-	Viterbi
VQ	-	Vector Quantization
Δ	-	First-order time derivatives
$\Delta\Delta$	-	Second-order time derivatives
a_{ij}	-	State transition probability from *i* to *j*
$b_j(k)$	-	Discrete probability Distribution at state j
$b_j(x)$	-	Continuous probability density function (pdf)
λ	-	HMM Model

α	-	Forward variable
β	-	Backward variable
$\gamma_t(i)$	-	Probability of being in state *i* at time *t*,
$\xi_t(i, j)$	-	Probability of being in state *i* at time *t*, and state *j*, at time $t + 1$
c_{jk}	-	Weight coefficient for the *m*th mixture component at state *j*
U_{jm}	-	Covariance matrix for the *m*th mixture component at state *j*
μ_{jm}	-	Mean vector for the *m*th mixture component at state *j*

LIST OF APPENDIXES

APPENDIX	TITLE	PAGE
A	Experimental Result	174
A.1.1	Test with Manual Segmentation & Unigram	174
A.1.2	Test with Manual Segmentation & Bi-gram	179
A.1.3	Test with Segmental K-Mean & Unigram	184
A.1.4	Test with Segmental K-Mean & Bi-gram	189
B	Source Code of the CDHMM Modeling (Program Functions)	197
B.1	Overview of CHDMM Modeling and Decoding Function	197
B.2	Description of CHDMM Training Implementation (System level)	199
B.2.1	BW Training	199
B.2.2	VB Training	200
B.2.3	BW+VB Training	201
C	Publications	203

CHAPTER 1

INTRODUCTION

1.1 Introduction

Speech is the primary mode of communication among humans and spoken language has become accepted as a natural method for human-machine interaction. Our ability to communicate with machines and computers, through keyboards, mice and other devices, is an order of magnitude slower and more cumbersome. In order to make this communication more user-friendly, speech input is an essential component. Besides that, natural speech contains a great deal of information that expressed by human. Even an illiterate or person with little knowledge about computer may use speech to operate computer. Many disabled people may use a computer with the help of speech input in case they are unable to type in keyboard or click in mouse with their hands. For normal people or experienced people in computer, they can utilize the speech input ability of a computer to significantly speedup documentation writing, email sending and other operation with computer. Another advantage of speech input is that it can be used for many situations when the hands are already used for important operations such as driving a car. Speech enabled dialing and GPS are two examples. With the development of Machine Translation techniques, another exciting application called Automatic Spoken Language Translation had emerged, which allow people from different countries all over the world to be able to freely communicate via speech without any professional translator.

One of the fundamental challenges of developing a spoken language system is the development of a speech recognition component. Research in speech recognition has been ongoing for approximately three decades. Much progress has been made during that time span. The technology started with very small vocabulary, speaker dependent, isolated word recognition systems. Today, the technology has been moved to large vocabulary systems, capable of recognizing from 20,000 to upwards of 100,000 words. The systems are now speaker independent, working out of the box for any speaker, and in some cases even speaker adaptive, learning the peculiarities of a person's speech over time. Isolated speech has long yielded to continuous speech in the research environment, and more recently, in the commercial marketplace as well, with the introduction of systems by IBM and Dragon. Error rates have been reduced dramatically.

There is a great potential for the application of the speech technology in Malaysia especially in the context of Malay speech. There is limited research on Malay speech recognition. Furthermore, the research of speech technology in Malaysia is still in its infancy stage. The development of the technology is limited to small vocabulary, isolated word application and lack of speaker-independency (Sh-Hussain 1993; Lim 2000; Hong 2004; Rubita *et al.* 2005). Beside that, such systems are still applying some conventional techniques of speech recognition. This constraints the recognition accuracy, robustness and adaptability of the systems.

This research aims to solve the above constraints of current Malay speech recognizers and provide a basis study and research on developing a medium vocabulary, speaker independent, Malay continuous speech recognition system. This work applies more robust pattern recognition techniques in Malay speech recognition. We use continuous density hidden Markov modeling (CDHMM), a more powerful modeling technique of speech as an alternative to existing techniques such as Discrete Markov Hidden Model (DHMM), Neural Network (NN) and Dynamic Time Warping (DTW). CDHMM which is more capable in modeling inter-speaker acoustic variability is expected to be able to relax the constraint of speaker-dependency. Although the CDHMM is used to solve the speaker-dependent task in this thesis and will provide a basis for solving speaker-independent task in the future. We also extend the existing Malay isolated word recognition system to Malay continuous speech recognition tasks by designing and developing word-based Malay connected word recognition system. This work also includes Malay phonetic segmentation and classification experiments as a preliminary research in using sub-word model as modeling unit which is needed in developing large vocabulary system This will provide basis on developing sub-word unit based Malay medium and large vocabulary continuous speech recognition system.

1.2 Problem Statement

The current Malay speech recognizers are limited by the following constraints: (1) speaker-dependency, (2) isolated words and (3) small vocabulary:

(1) ***Lack of speaker-independency***. Although there were researches on speaker-independent speech recognition in Malay speech domain, there is still room for improvement on speaker-independent recognition accuracy. Speaker-independent recognition is desirable to use a large number of speech parameters (or features). Thus, a modeling technique that can account for many parameters is needed. Due to the complexities introduced by freeing these constraints and the greater amount of training data available for speaker-independent recognition, efficient and automatic algorithms must exist for training and recognition of the model. Hidden Markov modeling (HMM) is powerful technique that capable for the robust modeling of speech. The currently used acoustic model in Malay speech recognizers are discrete Hidden Markov Model (DHMM). DHMM works considerably well in speaker-dependent tasks but the degradation of accuracy become apparent in speaker-independent tasks. Besides, DHMM suffers quantization error and cause accuracy

degradation. A more efficient modeling algorithm must be adapted to counter this limitation.

(2) ***Limitation to isolated word recognition.*** The current Malay speech recognition system is constrained to isolated word recognition tasks, mainly focused on isolated digit recognition. There is limited research on Malay continuous speech recognition system. There is a need to release this constraint by investigating ways to develop Malay continuous speech recognition system.

(3) ***Limitation to small vocabulary tasks.*** The current Malay speech recognizers are limited to small vocabulary recognition task, such as 10 Malay digit recognition. Word model is used as acoustic model. There is limited research on large or medium vocabulary task where the sub-word unit modeling is needed. Lack of research and study to be done on recognition using sub-word units such as phoneme, diphone and triphone. This constraint the recognition tasks to small vocabulary. There is need of applying sub-words as modeling unit in order to establish large and medium vocabulary Malay speech recognition system.

1.3 Objectives of the Research

The main objective of this study is to investigate ways to solve the speaker-dependence, isolated word, and small vocabulary constraints of current Malay speech recognizers and to provide preliminary study and research on developing Malay speaker-independent, medium vocabulary, continuous speech recognition system. To achieve the main objective, several sub-objectives are addressed in this thesis as following:

(1) To investigate the principle and architecture of HMM based statistical automatic speech recognition system.

(2) To apply continuous density HMM (CDHMM)) (Bahl et al. 1988, Poritz, & Richter 1986, Rabiner *et al.* 1985), which directly models the acoustic observation without VQ as an alternative to DHMM (Rabiner *et al.* 1983) in Malay speech recognition system. This is to eliminate the quantization errors caused by DHMM, thus increase the recognition accuracy. The CDHMM which is more capable of capturing inter-speaker acoustic variability and thus improve accuracy in speaker-independent (SI) recognition task compared to DHMM. The effectiveness of CDHMM is tested on speaker-dependent multi speaker task in this thesis as a basis for SI task in the future.

(3) To design and develop word-based Malay isolated word and continuous speech recognition system using CDHMM.

(4) To test CDHMM based Malay phoneme classification and segmentation on a medium vocabulary Malay continuous speech database as a preliminary study and research on sub-word unit modeling. The knowledge and experience gained is a basis for developing the sub-word unit based large and medium vocabulary Malay speech recognition system.

1.4 Scope of Research

The scope of task and the scope of approaches used in this thesis are defined as follows:

(1) The tasks to be solved in this thesis are as follow:

(a) The following speech recognition experiments were established:
- Isolated digit recognition
- Connected digit recognition
- Phoneme classification
- Phonetic segmentation

(b) All the experiments above are based on Malay speech domain.

(c) All the experiments above are speaker-dependent tasks.

(d) The phoneme classification and phonetic segmentation are based on medium vocabulary Malay speech database.

(2) The techniques and approaches used in solving the tasks are as follow:

(a) The HMM based statistical approach is used to develop the Malay automatic speech recognition system.

(b) Mel-frequency cepstral coefficient (MFCC), normalized energy, their first and second order derivatives (Delta and Delta-delta) (Furui 1986; Furui 1981) are used for feature extraction.

(c) Left-to-right continuous density hidden Markov model (CDHMM) with Gaussian mixture densities (Rabiner *et al.*1985; Juang *et al.* 1985) is used for acoustic modeling.

(d) Word model is used in isolated and connected digit recognition and sub-word model (phoneme) is used in phoneme classification and phonetic segmentation.

(e) The training algorithms used in the tasks are listed as follows:
- For isolated digit recognition, Baum-Welch algorithm (Rabiner 1989) and Viterbi algorithm algorithm (Rabiner *et al.* 1985; Juang *et al.* 1985) are used for training the

word models and comparison of their effect on recognition accuracy is made.

- For connected digit recognition, manual segmentation and segmental K-mean training strategies for continuous speech Rabiner (Rabiner *et al.* 1986a; Rabiner *et al.* 1986b) were used and comparison of their effect on recognition accuracy is made. Viterbi training algorithm is used for model re-estimation.
- For phoneme classification and phonetic segmentation, Viterbi training algorithm is used for training the phoneme models.

(f) Manually estimated bi-gram and calculated unigram (Jelinek 1991) value is used for language modeling for connected digit recognition.

(g) Viterbi full search algorithm (Viterbi 1967; Rabiner 1989) is used for decoding and Viterbi forced alignment is used for phonetic segmentation.

(h) Incorporating Malay phonetic knowledge in sub-word HMM based Malay speech recognition. This is done by identifying a Malay phone set, which well characterize Malay speech, to model.

(i) Effect on recognition and segmentation accuracy is investigated by varying number of Gaussian mixture components and using different combination of features.

1.5 Outline of Thesis

Chapter 2 describes literature review on the field of speech recognition. Various speech recognition tasks are described: speaker-dependent vs speaker-independent, isolated words vs continuous speech, small vocabulary vs large vocabulary. The constraint and difficulties of each task is discussed. Then, the current profound speech recognizers and their performance are presented. The current Malay speech recognizers are reviewed. From the review, their limitation and constraints are identified: speaker-dependent, isolated words, small vocabulary. Based on the review, derived the objective of the thesis to solve these constraints and provide a basis in developing speaker-independent Malay large vocabulary continuous speech recognition system. The different approaches applied to speech recognition are described. Next described are the various classification and modeling techniques in speech recognition. Their relative strengths and weaknesses are identified. From the review of different approach and techniques, the most suitable one are adapted, thus the scope of developments are identified.

Chapter 3 describes an overview of statistical speech recognition system and the use of HMM as statistical modeling of speech. The principle and architecture of the statistical speech recognition system are described. Mel-Frequency Cepstral Coefficient (MFCC) as acoustic front end processing is presented next. The theoretical foundation of Hidden Markov Modeling is discussed. The strength of HMM as applied to speech recognition is discussed in details. The various elements in HMM modeling such as re-estimation algorithm, model topology, observation probabilities distribution, and knowledge source representations are described in details. The variations of each element are also presented.

Chapter 4 describes the CDHMM based isolated words recognition system developed in this research. The performance of the system, evaluated on Malay isolated digit recognition task is presented and discussed.

Chapter 5 describes the CDHMM based connected words recognition system designed and developed in this research. The performance of the system, evaluated on Malay connected digit recognition task is presented and discussed.

Chapter 6 describes the HMM based phonetic classification and segmentation. A series of experiments are carried out, based on Malay continuous speech database to examine various elements of phonetic classification and segmentation.

Chapter 7, the final chapter, summaries the research findings. This chapter also identifies some problems of the techniques used in this research. Some suggestions for future work which might be useful for further development and improvement to the developed techniques.

1.6 Contribution of the Thesis

The current Malay speech recognizers are still constrained by speaker-independent small vocabulary isolated word recognition. The research findings provide a basis to develop speaker-independent Malay large vocabulary continuous speech recognition (LVCSR). The major contributions are as follows:

- CDHMM has been applied successfully in Malay speech recognition. The use of CDHMM performs better than the DHMM, which is currently used in Malay speech recognizers. The high accuracy achieved by CDHMM in speaker-dependent multi speaker task proves its ability to model inter-speaker variability. This result encourages its implementation in speaker-independent task.

- The CDHMM based connected word recognition system has been designed and developed. The evaluation on Malay connected digit recognition task achieved reasonable good results. The architecture of the word-based connected speech recognition system provides a basis to develop sub-word unit based continuous speech recognizers in the future.

- Various experiments on phonetic classification and segmentation on were conducted. The evaluation was based on medium vocabulary Malay continuous speech database. This provides experience and knowledge in sub-word unit modeling, which will be a basis to develop the large vocabulary sub-word based recognizers.

- The Malay phonetic knowledge has been successfully incorporated in HMM based Malay speech modeling. An adequate phone set in characterizing Malay speech is identified to model.

REFRENCES

Abdullah Hassan. (1980). *Linguistik Am Untuk Guru Bahasa Malaysia*,: Penerbit Fajar Bakti Sdn. Bhd., Petaling Jaya, Selangor.

Adell, J., Bonafonte, A., Gomez, J. A., Castro, M. J. (2005). Comparative Study of Automatic Phone segmentation Methods for TTS. In *Proc ICASSP*.

Antal, M. (2004). Speaker Independent Phoneme Classification in Continuous Speech. *Studia Univ Babes-Bolyai, Informatica,* vol. XLIX, no 2.

Atal, B. S. & Schroeder, M. R. (1967). Predictive Coding of Speech Signals. *IEEE Conference on Speech Communication and Processing*, pp. 360–361.

Bahl, L. R. (1980). Further Results on the Recognition of a Continuously Read Natural Corpus. *In Proc ICASSP.*

Bahl, L. R. *et al.* (1978). Automatic Recognition of Continuously Spoken Sentences from a Finite State Grammar. *In Proc ICASSP*, pp. 418-421.

Bahl, L. R., Bakis, R., Cohen, P. S., Cole, A. G., Jelinek, F., Lewis, B. L., Mercer, R. L. (1981) Speech Recognition of Natural Text Read as Isolated Words. *In Proc IEEE International Conference on Acoustics, Speech, and Signal Processing.*

Bahl, L. R., Brown, P. F., de Souza, P. V., and Mercer, R. L. (1988). Speech Recognition With Continuous Parameter Hidden Markov Models. *In Proceedings of the International Conference on Acoustics, Speech and Signal Processing.* pp. 40-43.

Bahl, L. R., Brown, P. F., de Souza, P. V., Mercer, R. L. (1986). Maximum Mutual Information Estimation of Hidden Markov Model Parameters for Speech Recognition. *IEEE International conference on Acoustic, Speech, and Signal Processing.*:49-52.

Bahl, L. R., Jelinek, F., & Mercer, R. (1983). A Maximum Likelihood Approach to Continuous Speech Recognition. *IEEE Trans. On Pattern Analysis and Machine Intelligence.* Vol. PAMI-5, pp. 179-190.

Baker, J. K. (1975). The DRAGON System-An Overview. *IEEE Transactions on Acoustic, Speech, and Signal Processing.* ASSP-23(1):24-29.

Baum, L. E. (1972). An inequality and associated maximization techniques in statistical estimation for probabilistic functions of Markov processes. *Inequalities*, 3: 1-8.

Baum, L. E., Petrie, T., Soules, G., and Weiss, N. (1970), A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains *The Annals of Mathematical Statistics,* 41: 164-171.

Becchetti C., Ricotti L. P. (2002) *Speech Recognition Theory and C++ Implementation.* West Sussex: John Wiley & Sons Ltd; 212-240.

Bilmes, J. and Zweig, G. (2002). The Graphical Models Toolkit: An Open Source Software System for Ppeech and Time-series Processing. *Proc. of ICASSP*, 4, pp. 3916–3919.

Brown, P. (1987). *The Acoustic Modeling Problem in Automatic Speech Recognition.* Ph.D. Thesis. Carnegie Mellon University.

Brugnara, F., Falavigna, D. & Omologo, M. (1992). A HMM-Based System for Automatic Segmentation and Labeling of Speech. *In Proc of ICSLP'92*, pp. 803-806.

Chen, S., Eide, E., Gales M. J., Ramesh, F,. Gopinath, A., Kanevsky, D., Olsen, P. (1999). Recent Improvements to IBM's Speech Recognition System for Automatic Transcription of Broadcast News, *Proceedings of the DARPA Broadcast News Workshop*, pp 89-94.

Chengalvarayan, R., Deng, L. (1998). Speech Trajectory Discrimination using the Minimum Classification Error Learning. IEEE Trans on Speech and Audio Processing. Vol. 6. No. 6 pp. 505-515.

Choukri, K., Chollet, G. (1986). Adaptation of Automatic Speech Recognizers to New Speakers using Canonical Correlation Techniques. *Computer Speech and Language.* Vol 1: 95-107.

Chow, Y. L., Dunham, M. O., Kimball, O. A., Krasner, M. A., Kubala, G. F., Makhoul, J., Roucos, S., Schwartz, R.M. (1987) BYBLOS: The BBN Continuous Speech Recognition System. *IEEE International conference on Acoustic, Speech, and Signal Processing,* pp. 89-92.

Cole, R. A. (1986a). Phonetic classification in New Generation Speech Recognition Systems. *In Proc of Speech Tech 86:* 43-46.

Cole, R. A., Philips, M., Brennan, B., Chigier, B. (1986b). The CMU Phonetic Classification System. *In Proc IEEE International Conference on Acoustics, Speech, and Signal Processing.*

Cole, R. A., Stern, R. M., Philips, M. S., Brill, S. M., Specker, P., Pilant, A. P. (1983). Feature based Speaker Independent Recognition of English Letters. *In Proc IEEE International Conference on Acoustics, Speech, and Signal Processing.*

Cox. S., Brady, R., & Jackson, P. (1998). Techniques for Accurate Automatic Annotation of Speech Waveforms. *In Proc of ICSLP'98*, vol. 5, pp. 1947-1950.

Davis, S. & Mermelsten, P., (1980). Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuous Spoken Sentences. *IEEE Trans. Acoustics, Speech and Signal Processing*, 28 (4), pp.357-366.

Deng, L. & O'Shaughnessy, D. (2003). *Speech Processing – A Dynamic and Optimization-Oriented Approach***,** Marcel Dekker, Inc.

Derouault, A. (1987). Context-dependent Phonetic Markov Models for large Vocabulary Speech Recognition. *In Proc ICASSP.* pp. 360-363.

Devillers, L. & Dugast, C. (1993). Combination of Training Criterion to Improve Continuous Speech Recognition. *In Proc European Conference on Speech Communication and Technology.* pp. 2211-2214.

Epraim, Y., & Rabiner, L. R. (1988). A Quantitative Assessment of the Relative Speaker Discriminating Properties of Phonemes. *IEEE International conference on Acoustic, Speech, and Signal Processing.* Vol. 1. 133-136.

Forney, G. D. (1973). The Viterbi algorithm. *Proc. IEEE*, vol. 61, pp. 268-278.

Furui, S. (1981). Cepstral Analysis Techniques for Automatic Speaker Verification. *IEEE Trans on ASSP* 29(2): 254-272.

Furui, S. (1986) Speaker Independent Isolated Word Recognition using Dynamic Features of the Speech Spectrum. *IEEE Trans on Acoustics, Speech and Signal Processing*, Vol. 34, No. 1, pp.52-59.

Ganapathiraju, A. (2002). *Support Vector Machine for Speech Recognition.* Ph.D Thesis. Carnegie Mellon University.

Gauvain, J. L., Lamel, L., Adda, G., and Jardin, M. (1999). The Limsi 1998 Hub-4e Transcription system. In *Proc. of the DARPA Broadcast News Workshop*, pp 99-104.

Gupta, V. N., Lennig, M., Mermelstein, P. (1987). Integration of Acoustic Information in a Large Vocabulary Word Recognizer. *In Proc ICASSP*, pp. 697-700.

Hermansky, H. (1990). Perceptual Linear Predictive (PLP) Analysis of Speech. *Journal of Acoust. Soc. Am.*, pp.1738-1752.

Hong, K. S. (2004). *The Design and Development of Educational Software on Automatic Speech Recognition.* Master Thesis. Universiti Teknologi Malaysia.

Huang, X. and Jack, M. A. (1989). Unified Techniques for Vector Quantization and Hidden Markov Models Using Semi- continuous Models. *In Proceedings of the International Conference on Acoustics, Speech und Signal Processing,* pp. 639-642.

Huang, X., Alleva, F., Hon, H. W., Hwang, M., Lee, K. F., Rosenfeld, R. (1993). The SPHINX-II Speech Recognition System: An Overview. *Computer Speech and Language,* vol. 2:137-148.

Hwang, M. Y. (1993). *Sub-phonetic acoustic modeling for speaker-independent continuous speech recognition.* Ph.D. Thesis. Carnegie Mellon University.

IBM speech recognition group. (1985) A Real-Time, Isolated-Word, Speech Recognition System for Dictation Transcription. *IEEE International conference on Acoustic, Speech, and Signal Processing.*

Jelinek, F. (1976). Continuous Speech Recognition by Statistical Methods. *Proc of the IEEE* 64(4):532-556.

Jelinek, F., (1991). Up from trigrams! The Struggle for Improved Language Models. *Proc. of Eurospeech,* pp.1037-1040.

Juang, B. H. & Rabiner, L. R. (1991). Hidden Markov Models for Speech Recognition. *Technometrics*, vol. 33, no 3, pp. 251-272, 1991.

Juang, B. H. (1985). Maximum Likelihood Estimation for Mixture Multivariate Stochastic Observations of Markov Chains. *AT&T Technical Journal*. 64. 1235-1249.

Juang, B. H., and Rabiner, L. R. (1990). The Segmental k-Means Algorithm for Estimating Parameters of Hidden Markov Models, *IEEE Transactions on Acoustics. Speech and Signal Processing. 38*: 1639-1641.

Juang, B. H., Rabiner, L. R., Levinson, S. E. & Sondhi, M. M. (1985). Recent Developments in the Application of Hidden Markov Models to Speaker-Independent Isolated Word Recognition, *In Proceedings of the ICASSP*, pp. 9-12.

Juang. B. H., and Rabiner, L. R. (1985). Mixture Autoregressive Hidden Markov Models for Speech Signals. *IEEE Transactions on Acoustics. Speech and Signal Processing,* 33, 1404-1413.

Kim, Y. J., Conkie, A. (2002). Automatic Segmentation Combining an HMM-based approach and Spectral Boundary Correction. *Proceedings of ICSLP.*

Lamere, P. *et al.* (2003). *Design of the CMU Sphinx-4 Decoder*. MITSUBISHI Electric Research Lab.

Lee, K. F. (1986). Incremental Network Generation in Word Recognition. *In Proc IEEE International Conference on Acoustics, Speech, and Signal Processing.*

Lee, K. F. (1988). *Large-Vocabulary Speaker-Independent Continuous Speech Recognition: The SPHINX System.* Ph.D. Thesis. Carnegie Mellon University.

Lee. K. F., Hon, H. W., and Reddy, R. (1990). An Overview of the SPHINX Speech Recognition System. *IEEE Transactions on Acoustics, Speech and Signal Processing*: vol. 38:35-45.

Lee. K. F., Hon. H., (1989). Speaker Independent Phone Recognition Using Hidden Markov Model, *IEEE Transactions on Acoustics Speech and Signal Processing*, pp 1641-1648,

Levinson, S. E., Rabiner, L. R Sondhi, M. M., (1983). Speaker Independent Isolated Digit Recognition Using Hidden Markov Model, *In Proceedings of ICASSP*, pp.1049-1052.

Levinson, S. E., Rabiner, L. R., Rosenberg, A. E., Wilpon, J. G. (1979). Interactive Clustering Techniques for Selecting Speaker-Independent Reference Templates for Isolated Word Recognition. *In IEEE Transactions on Acoustics, Speech, and Signal Processing ASSP*-27(2):134-141.

Levinson, S. E., Rosenberg, A. E., Flanagan, J. L. (1977). Evaluation of a Word Recognition System Using Syntax Analysis. *In Proc IEEE International Conference on Acoustics, Speech, and Signal Processing.*

Li, X. L. (2005). *Real Time Speaker Independent Large Vocabulary Continuous Speech Recognition.* Ph.D Thesis. University of Missouri.

Lim, S. C. (2000). *Isolated Word Speech Recognition Using Hidden Markov Models.* Bachelor Thesis. Universiti Teknologi Malaysia. 1-101.

Linde, Y., Buzo, A., Gray, R. M. (1980). An Algorithm for Vector Quatizer Design. IEEE Trans on Communication. 28(1): 84-95.

Liporace, L. A. (1982). Maximum Likelihood Estimation for Multivariate Observations of Markov Sources. IEEE Trans. Information Theory, vol. IT-28, no. 5, pp. 729-734.

Lippmann, R. P., Martin, E. A., Paul, D. P. (1987). Multi-Style Training for Robust Isolated Word Speech Recognition. *In Proc ICASSP.* pp. 705-8.

Lippmann, R. P., Martin, E. R., Paul, D. P. (1987). Multi Style Training for Robust Isolated Word Speech Recognition. *In ICASSP* pp. 705-8.

Ljolje, A., Hirschberg, J., van Santen, J. P. H. (1997). Automatic Speech Segmentation for Concateative Inventory Selection. In *Progress in Speech Synthesis*, van Santen, J. P. H, Springer-Verlag, New York.

Ljolje, A., Riley, M. D. (1991). Automatic Segmentation and Labeling of Speech. *In Proc of ICASSP*, pp 473-476.

Lowerre, B. T. (1977). Dynamic Speaker Adaptation in the Harpy Speech Recognition System. *In Proc IEEE International Conference on Acoustics, Speech, and Signal Processing.*

Merve R. van der. (1997). *Variations on Statistical Phoneme Recognition- A Hybrid Approach.* Master Thesis. University of Stellenbosch.

Ney, H. & Ortmanns, S. (1999). Dynamic Programming Search for Continuous Speech Recognition. *IEEE Signal Processing Magazine*, 16 (5), pp.64-83.

Ney, H. (1984). The Use of a One-Stage Dynamic Programming Algorithm for Connected Word Recognition. *IEEE Trans. on Acoustics, Speech, and Signal Processing,* 32(2):263-271.

Ney, H. (1987). Dynamic Programming Speech Recognition using Context Free Grammars. *In Proc ICASSP*, pp. 231-324.

Ney, H., Mergel, D., Noll, A., Paseler, A. (1987). A Data Driven Organization of the Dynamic Programming Beam Search for Continuous Speech Recognition. *In ICASSP*, pp. 833-836.

Ney, H., Noll. A. (1988). Phoneme Modeling using Continuous Mixture Densities. *Proceedings of the ICASSP*, pp. 437-440.

Nik Safiah Karim, Farid M. Onn, Hashim Haji Musa, and Abdul Hamid Mahmood. (1995). *Tatabahasa Dewan. New Ed.* Dewan Bahasa dan Pustaka, Kuala Lumpur.

Normadin, Y., Cardin, R., Mori, R. De. (1994). High Performance Connected Digit Recognition Using Maximum Mutual Information Estimation. *IEEE Trans. On Speech and Audio Processing.* 2(2), 299-311.

Ostendorf, M. and Roukos, S. (1989). A Stochastic Segment Model for Phoneme-based Continuous Speech Recognition. *IEEE Trans. Acoustics, Speech and Signal Processing*, 37 (1), pp.1957-1869.

Ostendorf, M. *et al*. (1996). From HMM's to Segment Models: A Unified View of Stochastic Modeling for Speech Recognition. *IEEE Transactions on Speech and Audio Processing*, vol. 4, no. 5, pp. 360-378.

Paul, D. B., Martin, E. A. (1988). Speaker Stress-Resistant Continuous Speech Recognition. *In Proc IEEE International Conference on Acoustics, Speech, and Signal Processing.*

Pellom, B. L., and Hansen, J. H. L. (1997). Automatic Segmentation and Labeling of Speech Recorded in Unknown Noisy Channel Enviroments. *In Proc of 1997*

ESCA-NATO Workshop on Robust Speech Recognition for Unknown Communication Channels. pp. 167-170.

Philip, C. & Moreno, P. (1999). On the Use of Support Vector Machines for Phonetic Classification. *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*.

Poritz, A. B., & Richter, A. G. (1986). On Hidden Markov Models in Isolated Word Recognition, *In Proceedings of the International Conference on Acoustics, Speech and Signal Processing, New York: IEEE*, pp. 705-708.

Rabiner, L. and Juang, B. H. (1993). *Fundamentals of Speech Recognition.* Englewood Cliffs, N.J.: Prentice Hall. pp.42-481.

Rabiner, L. R. (1989). A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE.* 77(2): 257 –286.

Rabiner, L. R., Juang, B. H., Levinson, S. E., & Sondhi, M. M. (1985). Recognition of Isolated Digits Using Hidden Markov Models with Continuous Mixture Densities. *AT&T Technical Journal,* 64:1211-1222.

Rabiner, L. R., Levinson, S. E., and Sondhi, M. M. (1983). On the Application of Vector Quantization and Hidden Markov Models to Speaker-Independent Isolated Word Recognition. *Bell System Technical Journal,* 62:1075-1105.

Rabiner, L. R., Levinson, S. E., Rosenberg, A. E., Wilpon, J. G. (1979). Speaker-Independent Recognition of Isolated Word Using Clustering Techniques. *In IEEE Transactions on Acoustics, Speech, and Signal Processing ASSP*-27(4): 336-349.

Rabiner, L. R., Wilpon J. G., Juang B. H. (1986b). A Continuous Training Procedure for Connected Digit Recognition. *In Proc of ICASSP.*

Rabiner, L. R., Wilpon, J. G., Soong, F. K. (1989) High performance Connected Digit Recognition Using Hidden Markov Models. *IEEE International conference on Acoustic, Speech, and Signal Processing*.

Rabiner. L. R. and Juang, B. H. (1986). An Introduction to Hidden Markov Models. *IEEE ASSP Magazine* 3(1):4-16.

Rabiner. L. R., Wilpon, J. G., and Juang, B. H. (1986a). A Segmental k-Means Training Procedure for Connected Word Recognition. *AT&T Technical Journal.* 65, :21-31.

Reddy, D. and Zue, V. (1983).Recognizing Continuous Speech Remains an Illusive Goal. *IEEE Spectrum*, pp. 84–87.

Rosenberg, A. E., Rabiner, L. R., Wilpon, J., Kahn, D. (1983). Demisyllable-Based Isolated Word Recognition System. *In IEEE Transactions on Acoustics, Speech, and Signal Processing ASSP*-31(3): 713-726.

Rubita Sudirman, Sh-Hussain Salleh, Ting, C. M. (2005) NN Speech Recognition Utilizing Aligned DTW Local Distance Scores. *9th International Conference on Mechatronics Technology*.

Russell, M. J. and Moore, R. K. (1985). Explicit Modeling of State Occupancy in Hidden Markov Models for Automatic Speech Recognition. *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, pp. 5-8.

Sakoe, H. and Chiba, S. (1978). Dynamic Programming Algorithm Optimization for Spoken Word Recognition. *IEEE Trans. on Acoustics, Speech, and Signal Processing,* 26(1): 43-49.

Schmidt, M. & Gish, H. (1996) Speaker Identification Via Support Vector Classifiers, *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, pp. 105-108.

Sh-Hussain Salleh. (1993). *A Comparative Study of The Traditional Classifier and the Connectionist Model for Speaker Dependent Speech Recognition System.* Master Thesis. Universiti Teknologi Malaysia.

Shikano, K., Lee. K, Reddy, D. R. (1986). Speaker Adaptation Through Vector Quantization. *In Proc IEEE International Conference on Acoustics, Speech, and Signal Processing.*

Tebelskis, J. (1995). S*peech Recognition using Neural Networks*, Ph.D. Dissertation, Carnegie Mellon University.

Thompson, H. S., Laver, J. D. (1987). The Alvey Speech Demonstrator – Architecture, Methodology, and Progress to Date. *In Proc of Speech Tech.*

Ting, H. N. (2002). *Speech Analysis and Classification Using Neural Networks for Computer-Based Malay Speech Therapy.* Master Thesis. Universiti Teknologi Malaysia. 1-146.

Toledano, D. T. Rodríguez, M. A. and Escalada, J. G. (1998) Trying to mimic human segmentation of speech using HMM and fuzzy logic post-correction rules. In

Proceedings of the 3rd ESCA/COCOSDA International Workshop on Speech Synthesis, pp. 207–212.

Toledano, D. T., A. Hernandez Gomez, and Luis Villarrubia Grande. (2003) Automatic Phone Segmentation. *IEEE Transactions on Speech and Audio Processing*, pp. 617-625.

Vergin, R., O'Shaughnessy, D. Farhat, (1999). A Generalized Mel-frequency Cepstral Coefficients for Large Vocabulary Speaker Independent Continuous-Speech Recognition. *IEEE Transactions on Speech and Audio Processing*, 7(5):525 –532.

Viterbi, A. J. (1967). Error Bounds for Convolutional Codes and an Asymptotically Optimal decoding algorithm. *IEEE Trans. Information. Theory*, vol. IT-13, pp. 260-269.

Waibel. A. H. (1986). *Prosody and Speech Recognition.* Ph.D. Thesis. Computer Science Department, Carnegie Mellon University.

Wang, J. C., Wang, J. F. and Weng Y. S. (2000) Chip Design of Mel Frequency Cepstral Coefficients for Speech Recognition. *Proceedings of International Conference on Acoustics, Speech, and Signal Processing,* vol. 6. 3658-3661.

Wightman, C. W. and Talkin, D. T. (1997). The Aligner: Text-to-Speech Alignment Using Markov Models. In *Progress in Speech Synthesis*, Springer-Verlag, New York.

Wong, E. and Sridharan, S. (2001) Comparison of Linear Prediction Cepstrum Coefficients and Mel-Frequency Cepstrum Coefficients for Language

Identification. *Proceedings of 2001 International Symposium on Intelligent Multimedia, Video and Speech Processing*.

Wong, P. H. W. *et al.* (1998). Reducing Computational Complexity of Dynamic Time Warping Based Isolated Word Recognition with Time Scale Modification. *Proceedings of 1998 Fourth International Conference onSignal Processing Proceedings.* vol.1: 722 -725.

Young, S., Evermann, G., Kershaw, D., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V. and Woodland, P. (2002). T*he HTK Book (for HTK Version 3.2).* Microsoft Corporation and Cambridge University Engineering Department, England.