MALAY ARTICULATION SYSTEM FOR EARLY SCREENING DIAGNOSTIC USING HIDDEN MARKOV MODEL AND GENETIC ALGORITHM

MOHD NIZAM BIN MAZENAN

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Doctor of Philosophy (Biomedical Engineering)

Faculty of Biosciences and Medical Engineering
Universiti Teknologi Malaysia

FEBRUARY 2016

Dedicated to Papa and Mama,

my Sister (Mimi) and Brother (Agung)

Haziq and Maisarah

and my beloved family.

ACKNOWLEDGEMENT

First and foremost, I would like to thank my supervisor in this PhD study which is Dr. Tan Tian Swee and my co-supervisor Dr Nasrul. They have given me a lot of advice, chances, knowledge and guidance in completing this research study. They both always help me when I face problem in my research and study. To me, they are more like friends in which I am deeply grateful for their valuable guidance and continuously encourage me throughout my PhD journey.

I also wish to deliver my heartiest gratitude to my sensei Professor Hirofumi Nagashino and Associate Professor Akutagawa Masatake during my attachment in School of Health Sciences, Faculty of Medicine, The University of Tokushima, Japan. Nagashino sensei has guided me throughout the attachment in Tokudai especially in Medical problem and Akutagawa sensei has helped me in solving some technical problem especially in features and signal processing for my research studies. They both also help me regarding the placement and guide me in how to live in Japan.

Besides, I would like to thank to my amazing laboratory mate and they are Leong Kah Meng, Lum Kim Yun, Lau Chee Yong, Lim Yee Chea, Rina Salleh, and Gan Hong Seng. They have given me a plenty mental support and advice during my laboratory time.

Last but not least, I would like to thank my parents (Papa and Mama), my sister (Mimi), my brother (Agung), Abang Bad, Haziq, Maisarah and my special one Sarah Samson Soh as they have provided me precious moral and mental support to make me travel this far for all this years. They are always the one I want to make them proud. Thank you.

ABSTRACT

Speech recognition is an important technology and can be used as a great aid for individuals with sight or hearing disabilities today. There are extensive research interest and development in this area for over the past decades. However, the prospect in Malaysia regarding the usage and exposure is still immature even though there is demand from the medical and healthcare sector. The aim of this research is to assess the quality and the impact of using computerized method for early screening of speech articulation disorder among Malaysian such as the omission, substitution, addition and distortion in their speech. In this study, the statistical probabilistic approach using Hidden Markov Model (HMM) has been adopted with newly designed Malay corpus for articulation disorder case following the SAMPA and IPA guidelines. Improvement is made at the front-end processing for feature vector selection by applying the silence region calibration algorithm for start and end point detection. The classifier had also been modified significantly by incorporating Viterbi search with Genetic Algorithm (GA) to obtain high accuracy in recognition result and for lexical unit classification. The results were evaluated by following National Institute of Standards and Technology (NIST) benchmarking. Based on the test, it shows that the recognition accuracy has been improved by 30% to 40% using Genetic Algorithm technique compared with conventional technique. A new corpus had been built with verification and justification from the medical expert in this study. In conclusion, computerized method for early screening can ease human effort in tackling speech disorders and the proposed Genetic Algorithm technique has been proven to improve the recognition performance in terms of search and classification task.

ABSTRAK

Pada masa kini, aplikasi pengecaman pertuturan sudah menjadi tidak asing lagi sebagai satu teknologi yang mampu membantu individu kurang upaya. Terdapat kajian yang meluas dan pembangunan yang pesat telah dijalankan sejak beberapa dekad yang lalu. Namun begitu, bagi prospek kajian di Malaysia, ianya masih baru dan belum matang dari aspek penggunaan dan pendedahan walaupun terdapat permintaan yang tinggi daripada sektor perubatan dan penjagaan kesihatan. Tujuan utama kajian ini adalah untuk menilai kualiti dan kesan penggunaan kaedah komputeran bagi tujuan peringkat awal diagnosis bagi mengatasi masalah gangguan pertuturan artikulasi untuk konteks di Malaysia. Di dalam kajian ini, kaedah kebarangkalian statistikal menggunakan Hidden Markov Model (HMM) telah diadaptasi dengan reka bentuk corpus bahasa Melayu baru untuk permasalahan artikulasi berpandukan skrip SAMPA dan garis panduan yang telah ditetapkan oleh IPA bersama-sama pengesahan daripada terapis pertuturan di Hospital. Penambahbaikan telah dilakukan bermula dari peringkat pemprosesan bahagian depan bagi pemilihan vektor ciri dengan menerapkan algoritma kalibrasi kawasan tanpa isyarat suara untuk pengesanan titik permulaan dan akhir. Pengkelas yang digunakan turut ditambah baik dengan menggunakan teknik penghibridan carian Viterbi dengan genetik algoritma (GA) bagi memperoleh ketepatan yang tinggi dalam keputusan pengecaman dan juga proses klasifikasi unit leksikal. Keputusan kajian dinilai berpandukan tanda aras oleh NIST. Keputusan kajian menunjukkan ketepatan pengecaman telah meningkat kira-kira 30% hingga 40% dengan menggunakan teknik hibridisasi Viterbi dan genetik algoritma berbanding teknik konvensional. Kesimpulannya, pembentukan corpus baru telah berjaya dibina dengan pengesahan dan justifikasi oleh pakar perubatan. Oleh yang demikian, penggunaan komputer untuk pemeriksaan awal, dapat meringankan usaha manusia dalam diagnosis gangguan pertuturan dan teknik genetik algoritma yang diperkenalkan berjaya meningkatkan prestasi pengecaman dalam aspek tugas pencarian dan pengkelasan.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AAC	-	Alternatives And Augmentative Communication
AI	-	Artificial Intelligence
ASR	-	Automatic Speech Recognition
FE	-	Feature Extraction
FFT	-	Fast Fourier Transform
FSM	-	Finite State Machine
GA	-	Genetic Algorithm
GMM	-	Gaussian Mixture Model
HMM	-	Hidden Markov Model
HSA	-	Hospital Sultanah Aminah
IPA	-	International Phonetic Association
SAMPA	-	Speech Assessment Methods Phonetic Alphabet
SKPK	-	Sekolah Kebangsaan Pendidikan Khas
STE	-	Short-Time Energy
KKM	-	Kementerian Kesihatan Malaysia
KPM	-	Kementerian Pengajian Malaysia
LPC	-	Linear Predictive Coding
MFCC	-	Mel-Frequency Cepstral Coefficients
NIST	-	National Institute of Standards and Technology
OOV	-	Out-of-Vocabulary
SLP	-	Speech Language Pathologists
ST	-	Speech Therapist
STT	-	Speech-to-Text
WER	-	Word Error Rate
WHO	-	World Health Organization
ZCR	-	Zero Crossing Rate

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Individuals who are diagnosed with speech disorder have problems creating or forming speech sounds to communicate with others. This may lead to several speech impairments such as articulation disorders, fluency disorders and voice disorders. Statistic by World Health Organization (WHO) shows that, speech and language disorder affects at least 3.5% of human population communication skills. In Malaysia, its affect 5-10% of children with manifest speech and language problems totaling about 10,223 children were reported to have learning disabilities (KPM 2002, Woo and Teoh 2007, Raflesia 2008). This research is written to highlight and focus more on the main issue in medical speech that is related to early screening of articulation disorder case. Previous research shows that, early speech therapy intervention is critically important as well as other developmental concerns as a part of knowing what's "normal" and what's “not” in speech and language development. This can help parents or guidance especially Speech Therapist (ST) to figure out the next concern on right schedule and to prevent possible future problems (Fey 1986, Nelson 1989, Warren and Yoder 1996, Trevino-Zimmerman 2008, Navarro-Newball, Loaiza et al. 2014, Rubin, Kurniawan et al. 2014, Sparreboom, Langereis et al. 2015).

Articulation disorders are errors in the production of speech sounds that may be related to anatomical or physiological limitations in the skeletal, muscular, or neuromuscular support for speech production (Gargiulo 2006) which may also interfere with intelligibility (Association 1993). These disorders include omissions (bo for boat), substitutions (wabbit for rabbit), additions (animamal for animal) and distortions (shlip for sip) (Bauman-Waengler 2012). The degrees of disorders vary from mild impairments like pronunciation errors to more severe ones including hearing-loss, aphasia, and craniofacial anomalies. Some children or even adults, have trouble saying certain words, sounds or sentences where others have trouble understanding what they are trying to say. However, there are special therapies such as Speech Therapist (ST) and Speech Language Pathology (SLP) that can help these patients to go through the early screening diagnosis, assessment, and therapeutic assistance to help this patient to improve their speech.

The advanced research in speech and signal processing technology has open massive development of computer-assisted methods for speech therapy or called assistive device (AD). Speech-to-Text recognition system can be applied to help people who have speech impairment. It is to assist the speech therapist in diagnosing the patient. This idea has been well mentioned where the Automatic Speech Recognition (ASR) systems may be used for therapeutic applications, assessment applications, report writing and as a mode of alternatives and augmentative communication (AAC) (Griffin, Wilson et al. 2000). AD has widely been used for much therapeutic assistance for ST and few research in this area in Malaysia include Malay speech therapy assistance tools (MSTAT) (Tan, Ariff et al. 2007) and Computer-based Malay speech analysis system (Liboh 2011). Communication disorder by nature is very complex and those existing diagnostic tool is still incomplete and imprecise which need a lot more improvement. The research will focus more on the design and the development of Malay articulation disorder early screening system to assist in speech communication by using speech recognition engine that meet specific criteria. Speech therapy early screening diagnostic has been believed to help ST by doing the quick assessment to verify the correct level of disorder of the speech disorder patient (Misman 2014).

Speech-to-Text or often called speech recognition is a translation of spoken words into text (Rabiner and Juang 1993). Speech signal processing includes the acquisition, manipulation, storage, transfer and output of human vocal utterances by a computer (Gurjar and Ladhake 2008). Speech feature extraction work in front end and play an important role in the conversion and extraction of useful meaning from speech. The extracted speech features are used as input for speech recognition.

Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems in manipulation of the collected speech vector while still describing the data with sufficient accuracy. To produce the right speech recognition system specifically as a speech diagnosis tool that covered articulation case, there are few criteria need to be met such as the design of the Malay corpus, segmentation of speech utterance to specific lexical unit, and the robust pattern matching algorithm to spot and classifies the error in the speech utterance sample. Malay corpus-based has been designed with the collaboration of ST at Hospital Sultanah Aminah (HSA) and Sekolah Kebangsaan Pendidikan Khas (SKPK) whereby the process include mapping the word list into table of phoneme and manner of articulation including the pictogram rule specified by ST worldwide (Bell and Lindamood 1991, Gambrell and Jawitz 1993). This method of designing corpus will be based on human sound production and specifically to cover the process on human speech. Another aspect is the segmentation method of the speech utterance smallest lexical unit where most of the process been applied using Hidden Markov Model (HMM)-based speech recognition (Rabiner 1989). Following this method such as wavelet spectrum method (Ziółko, Manandhar et al. 2006) and segment features approach (Rybach, Gollan et al. 2009) which is the process to analyze and validate a computational model for the speech feature. The main purpose for the segmentation in speech recognition is to identify the boundaries between words, syllables, or phonemes in spoken natural language. Since speech utterance will be recognized by phoneme classifier, this utterance need to be segmentized into the smallest lexical unit to achieve higher recognition output. Back-end classification is another important process in this research which include acoustic modeling, pronunciation dictionary and language model as the ASR components. The statistical methods based on HMM is to produce a model from the speech training data which later on be use to predicts the target values of the speech testing data (Suuny, David Peter et al. 2013) using best pattern matching algorithm as the classifier which is treated as a search problem in this research.

1.2 Background of the Problem

Computer-based Malay speech early diagnostic system is still new in Malaysia and most of the systems available now are in foreign language such as English (Ting, Yunus et al. 2003). Most of the diagnostic system is not using the corpus specifically designed and verified by expert in this area such as ST. There is no specific or standard formatting in creating the corpus whether it is for early screening or diagnosis, treatment or training. All current wordlist design used by ST (especially in HSA and SKPK) are using the old wordlist developed by Ministry of Health Malaysia (KKM 2014, Misman 2014) in which most of the words may not be accurate enough to be used. Therefore the development of the corpus will follow specific rule to address articulation disorder case based on verification by ST and SLP. This is important aspect as the main purposed is to overcome the problem regarding articulation disorder early screening diagnosis. Another important aspect that may lead to the improvement in development of high quality speech-to-text system which is the segmentation of the acoustic utterance selected from the large-scaled corpus of isolated words that been specifically designed to fit Malay language that cover wide range of articulation words. Thus, a robust segmentation is needed to handle this sample by considering the aspect of right FE technique and the improvement of silence region calibration algorithm (Bhattacharjee 2013, Shanthi Therese and Lingam 2013). The effect of ignoring this aspect may lead to problem where many short silence tags are unreliable and hence lead to merging with neighboring segments (Hain, Johnson et al. 1998) which may reduce the recognition accuracy. After the useful feature has been extracted, segmentized and model, the pattern matching algorithm will take place as the speech recognizer to define each model to the right place according to the artic sample utterance by the patient of speech disorder.

1.3 Problem Statement

i) ST is deals with speech disorder patient where the process of early screening diagnosis can be conducted by using computerized speech recognition system to provide a way to assess the speech problem of the patient (Glykas and Chytas 2004). Therefore, it is an issue here on how to create the right speech modeling that can recognizer speech disorder unknown utterance by achieving high accuracy of recognition speech.

ii) The way of representing the speech output and transcriptions result must be effective to show the right disorder categories with met additional criteria such as quicken the process of therapy screening session but maintain the accuracy result. This is the most important element in a speech therapy program but the most difficult for a therapist to carry out, due to time limits (Bälter, Engwall et al. 2005). Therefore computational time of the system, accuracy, and intelligibility is an issue here.

iii) The wordlist of the Malay speech corpus design must contain sufficient prosodic and acoustic variations with significant unit following the right design and rule for articulation disorder early diagnosis. But, the current Malay speech database design is lack of the verification from medical expert and some of the target selection word not following “stimulable” therapy word. Therefore, it is an issue to develop the robust corpus that is based on the interwoven ideas of developmental readiness, ease of learning and understanding and easier for the clinician to teach (Van Riper and Irwin 1958, Van Riper 1978, Hodson 2007).

iv) Research been done in Malaysia shows that, this traditional method (hearing and experience judgment) is still in use widely in Malaysia hospital or speech center. There are no wrong doing by manual technique but based on observation to the hospital or clinic that still using this technique, the method sometimes lack of accuracy, time consuming (Ai and Yunus 2007) and require high number of ST for each session (Saz, Yin et al. 2009). The common practice of this process going through, the "tool" that only been rely by the ST is their hearing and experience judgment. That would go on for an hour and sometimes it would lead to error elimination during the process (Secord, Boyce et al. 2007).

1.4 Objectives of the Research

This research study is aiming to solve related problems in building Malay early screening computerized system for articulation disorder diagnosis. Therefore the objectives are:

i) To develop high quality Malay articulation early screening computerized system (in term of providing higher speech recognition accuracy and robust corpus design) for articulation disorder diagnosis use as therapeutic assistance for ST.

ii) To design a new set of Malay SAMPA database that specifically for computer-readable script that accommodate Malay articulation pronunciation phoneme applied in speech-to-text system.

iii) To implement and improve acoustic phonetic segmentation in speech recognition boundaries identification between smallest lexical units in spoken natural language with silence region calibration algorithm.

iv) To improve Viterbi searching algorithm by hybridization with Genetic Algorithm (GA) which work in back-end phase that provide optimal solution when solving artic lexical unit selection task.

1.5 Scope of the Research

The current corpus from Kementerian Kesihatan Malaysia (KKM) is based on Malay language for speech disorder diagnosis by manual technique that contains 108 target words with no specific or standard formatting in creating the corpus whether it's for early screening or diagnosis, treatment or training. A new speech database is built specifically for articulation disorder early screening followed specific rule in qualitative approach and target selection in phonological intervention. The target word has been simplified to 76 words or equivalent to 36480 training sampling sum up by 80 x 76 x 6 which considering the total 80 patient with 6 times repetition for each word. The subject patient is from 2 main races in Malaysia (Malay and Chinese), with both male and female, ranging from normal and articulation disorder patient. The age coverage is starting from the age of 9 until 23 years old.

The statistical engine used in this research is based on HMM as the auto-segment of the recorded speech sample to cover three lexical units of phoneme, syllable and word based, and also work as probabilistic rule for the speech modeling phase. The speech recognition processing module includes the segmentation, silence region calibration, feature extractor (based on MFCC) and pattern matching engine. These methods are also includes Genetic Algorithm (GA), HMM probabilistic method and the hybridization of Viterbi search and GA. The evaluation of the recognition system was tested based on Word Error Rate (WER), Out-of-Vocabulary (OOV), %Correct and Recognition Rate accuracy which been used as objective evaluation of the recognition system based on NIST benchmarking. This can measure the recognition performance, vocabulary identification and intelligibility of the output speech utterances.

1.6 Significance of the Study

This study implements speech recognition technology with improvements of HMM as the speech engine, silence region calibration algorithm in acoustic segmentation phase and hybridization of Viterbi search with GA for the classification process. The articulation early screening case is often referred to the problem of producing speech sounds related to smallest lexical unit such as phoneme and syllable. therefore, the purposed of solving this case is by tackling the phonology dimension in term of articulatory phonetics (transmission of sounds) by following the linguistic guideline based on the table of placed and articulation manner for human speech production. It also essential to preoccupied sound as a system that carrying information where is the task is to identify and recognize the correct phonemes. The evaluations of the recognition system were tested based on values of objective functions obtained, accuracy of the recognition and comparison test. By doing so, the advantage and disadvantage of the improvement methods used are known. Besides, new algorithm of silence region calibration in segmentation phase and the hybridization of Viterbi search and GA in classification process can be maintained the advantageous of both method. This new algorithm is able to improve the performance and obtain good solution in lexical unit recognition process that definitely makes contribution in the area of speech recognition specifically for articulation disorder early screening diagnosis.

1.7 Research Methodology

The design of Malay Speech-to-Text (STT) system is divided into three steps as shown in Figure 1.1. The system starts with first step of Malay speech corpus design or database development with specific rule verify by ST. The second step is the front-end processing component that process sampled speech signal into parametric representation by FE. This process also includes the silence region calibration, segmentation of lexical unit, and pattern modeling of the training data. The pattern of phoneme will be the templates for the next phase that was pattern recognition. The final step or back-end processing will deal with the classifier and decoder component for the recognition of unknown speech signal by the patient or normal speaker. The whole system is also been evaluated in this phase based on objective evaluation and also comparison test.

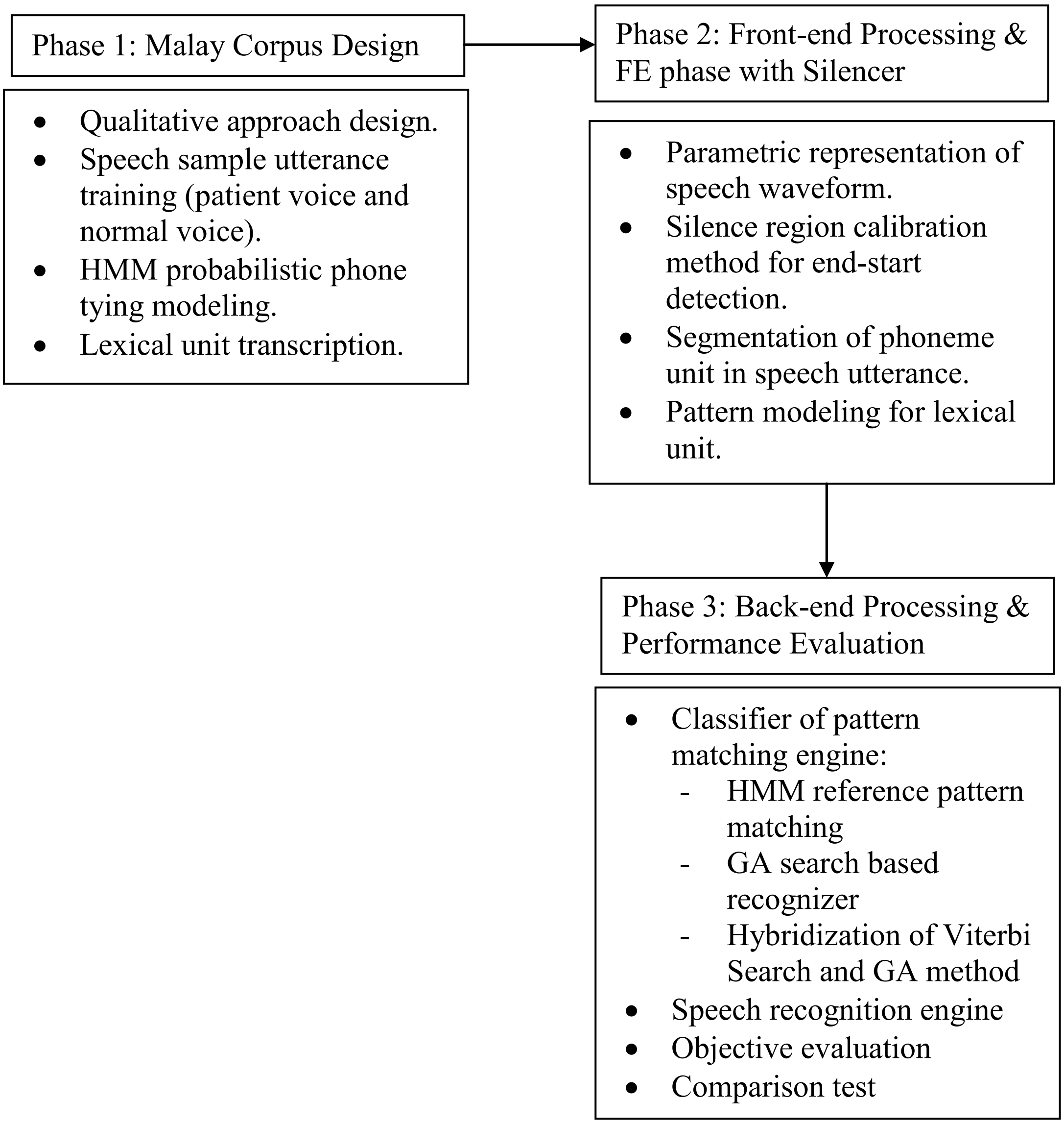

Figure 1.1. Mapping of procedures in this study

1.8 Thesis Layout

This thesis is divided into five chapters. Chapter 1 includes introduction, background of the problem, objectives of the research and scope of the thesis. The purpose is to show the initiative in using the proposed recognition method rather than current conventional or traditional method.

Chapter 2 will provide a literature review of this study. It includes reviews from previous study of the Text-to-Speech system, speech recognition and few methods in developing the recognizer machine. The focus is to develop the robust corpus for early screening diagnosis system of articulation speech disorder, the segmentation of smallest lexical unit method and the pattern matching in classifier phase. This chapter also included review of standard method in feature extraction, segmentation technique, silence region calibration, HMM recognizer engine, Viterbi search and GA for pattern matching.

Chapter 3 is the methodology used in this study. It describes the Malay early screening diagnostic system and the implementation specifically for articulation speech disorder. This chapter discussed the overall implementation, design, and the development of Malay speech corpus, cost function used and objective evaluation of the target cost. It also describes the procedure to implement the silence region calibration algorithm before segmentation phase for the front-end. The performance will be compared with previous method (without silencer and with silencer). Then the segmentation will be done by using the correct MFCC adjustment for the training sample. The performance evaluation methods are based on the accuracy percentage of word spotting. Another method discussed is regarding the procedure to implement HMM recognizer emphasizing the Viterbi search as a classifier for reference pattern matching. Then the procedure of hybridization with GA as a search based recognizer for the pattern matching and word spotting in decoder. The performance will be compared with conventional method.

Chapter 4 is the result and discussion section. It shows the result accuracy achieved by the recognizer and also describes the performance evaluation of all method including corpus design testing, smallest lexical unit segmentation, silence region calibration, FE adjustment, and the pattern matching in recognition phase. Their performance will be based on the evaluation of the recognition system was tested based on Word Error Rate (WER), Out-of-Vocabulary (OOV), %Correct and Recognition Rate accuracy which been used as objective evaluation of the recognition system. The entire test is based on NIST standard for ASR system.

Chapter 5 outlined the conclusion for the proposed technique and method which reflect the novelties of the research. This chapter will also include the future plan and the recommendation for further improvement of this study.

REFERENCES

Aarnio, T. (1999). *Speech recognition with hidden markov models in visual communication*. Master of Science Thesis, University of Turki.

Acero, A. (1995). *The role of phoneticians in speech technology*. BLOOTHOOFT, G.-HAZAN, V.-HUBER, D.

Ahadi, S., et al. (2003). An efficient front-end for automatic speech recognition. *Electronics, Circuits and Systems*, *2003. ICECS* 2003. Proceedings of the 2003 10th IEEE International Conference on, IEEE.

Ai, O. C. and J. Yunus (2007). Computer-based System to Assess Efficacy of Stuttering Therapy Techniques. *3rd Kuala Lumpur International Conference on Biomedical Engineering* 2006, Springer.

Aida-Zade, K., et al. (2006). Investigation of combined use of MFCC and LPC features in speech recognition systems. *World Academy of Science, Engineering and Technology* 19: 74-80.

Al-Haddad, S., et al. (2006). Automatic segmentation and labeling for Malay speech recognition. *WSEAS Transactions on Signal Processing*(9).

Al-Haddad, S., et al. (2009). Robust speech recognition using fusion techniques and adaptive filtering. *American Journal of Applied Sciences* 6(2): 290-295.

Al-Haddad, S. A. R., et al. (2007). Decision Fusion for Isolated Malay Digit Recognition Using Dynamic Time Warping (DTW) and Hidden Markov Model (HMM). *Research and Development, 2007. SCOReD* 2007. 5th Student Conference on, IEEE.

Al-Sulaiti, L. and E. S. Atwell (2006). The design of a corpus of contemporary Arabic. I*nternational Journal of Corpus Linguistics* 11(2): 135-171.

Ali, M. E. M., et al. (2007). Automatic Segmentation of Arabic Speech. *Workshop on Information Technology and Islamic Sciences*, Imam Mohammad Ben Saud University, Riyadh, March.

Alkulaibi, A., et al. (1997). Fast 3-level binary higher order statistics for simultaneous voiced/unvoiced and pitch detection of a speech signal. *Signal processing* 63(2): 133-140.

Angelini, B., et al. (1993). Automatic segmentation and labeling of english and italian speech databases. *Third European Conference on Speech Communication and Technology*.

Ankit Kumar, M. D. a. T. C. (2014). Continuous Hindi Speech Recognition using Monophone based Acoustic Modeling. *IJCA Proceedings on International Conference on Advances in Computer Engineering and Applications (ICACEA)*, Ghaziabad, UP, India IEEE.

Anusuya, M. and S. Katti (2011). Front end analysis of speech recognition: a review. *International Journal of Speech Technology* 14(2): 99-145.

Ariff, A. K., et al. (2005). Malay speaker recognition system based on discrete HMM. *Proceeding of 1st Conference on Computers, Communication, and Signal Processing, IEEE*.

Aronowitz, H., et al. (2004). Text independent speaker recognition using speaker dependent word spotting.

Arslan, L. M. (1999). Speaker transformation algorithm using segmental codebooks (STASC). *Speech Communication* 28(3): 211-226.

Association, A. S.-L.-H. (1993). *Definitions of communication disorders and variations*.

Association, I. P. (1999). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*, Cambridge University Press.

Association, I. P. (2005). *IPA Chart*, Creative Commons Attribution-Sharealike 3.0 Unported License. .

Atal, B. S. and L. Rabiner (1976). A pattern recognition approach to voiced-unvoiced-silence classification with applications to speech recognition. A*coustics, Speech and Signal Processing, IEEE Transactions* on 24(3): 201-212.

Axelsson, A. and E. Björhäll (2003). *Real time speech driven face animation*.

Aymen, M., et al. (2011). Hidden Markov Models for automatic speech recognition. Communications, Computing and Control Applications (CCCA), 2011 *International Conference on, IEEE*.

Bachu, R., et al. (2008). Separation of Voiced and Unvoiced using Zero crossing rate and Energy of the Speech Signal. *American Society for Engineering Education (ASEE) Zone Conference Proceedings*.

Bahl, L. R., et al. (1991). *Method and apparatus for the automatic determination of phonological rules as for a continuous speech recognition system*, Google Patents.

Baker, J. (1975). The DRAGON system--An overview. A*coustics, Speech and Signal Processing, IEEE Transactions* on 23(1): 24-29.

Bälter, O., O. Engwall, A.-M. Öster and H. Kjellström (2005). Wizard-of-Oz test of ARTUR: a computer-based speech training system with articulation correction. *Proceedings of the 7th International ACM SIGACCESS Conference on Computers and Accessibility, ACM*.

Balyan, A., et al. (2010). *Phonetic Segmentation Based On Hmm Of Hindi Speech*. The International Committee for the Co-ordination and Standardization of Speech Databases and Assessment Techniques.

Banerjee, P., et al. (2008). Application of triphone clustering in acoustic modeling for continuous speech recognition in Bengali. *Pattern Recognition, 2008. ICPR 2008*. 19th International Conference on, IEEE.

Basir, S. (1993). *An investigation of an object-oriented design implementation of a phoneme recognition system based on dynamic time warping algorithm*. Faculty of Electrical Engineering. Skudai, Johor Malaysia, Universiti Teknologi Malaysia. Masters.

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

Baum, L. E. and J. A. Eagon (1967). An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology. *Bull. Amer. Math. Soc* 73(3): 360-363.

Baum, L. E., et al. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The annals of mathematical statistics*: 164-171.

Bauman-Waengler, J. (2012). *Articulatory and phonological impairments: A clinical focus*, Pearson Higher Ed.

Bauman-Waengler, J. (2012). *e-Study Guide for: Introduction to Phonetics and Phonology : From Concepts to Transcription* by Jacqueline Bauman-Waengler, ISBN 9780205402878, Cram101.

Bayes, M. and M. Price (1763). An Essay towards solving a Problem in the Doctrine of Chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFRS. *Philosophical Transactions* (1683-1775): 370-418.

Bell, N. and P. Lindamood (1991). *Visualizing and verbalizing: For language comprehension and thinking*, Academy of Reading Publications Paso Robles, CA.

Bellman, R. (1956). Dynamic programming and Lagrange multipliers. *Proceedings of the National Academy of Sciences of the United States of America* 42(10): 767.

Benouareth, A., et al. (2006). Semi-continuous HMMs with explicit state duration applied to Arabic handwritten word recognition. *Tenth International Workshop on Frontiers in Handwriting Recognition*, Suvisoft.

Bharathi, C. and V. Shanthi (2012). Disorder Speech Clustering For Clinical Data Using Fuzzy C-Means Clustering and Comparison With SVM classification. *Indian Journal of Computer Science and Engineering (IJCSE)* 3(5).

Bhattacharjee, R. (2014). *Identification of Voice/Unvoiced/Silence regions of Speech*. Retrieved Jun, 2014, from http://iitg.vlab.co.in/?sub=59&brch=164&sim=613&cnt=1.

Bishop, C. M. (2006). *Pattern recognition and machine learning*, springer New York.

Bitar, N. N. and C. Y. Espy-Wilson (1996). Knowledge-based parameters for HMM speech recognition. *Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on, IEEE*.

Black, A. W. and K. A. Lenzo (2001). *Optimal data selection for unit selection synthesis*. 4th ISCA Tutorial and Research Workshop (ITRW) on Speech Synthesis.

Bleile, K. (2014). *The Manual of Speech Sound Disorders: A Book for Students and Clinicians*, Cengage Learning.

Bogdanova, A. M., & Gjorgjevikj, D. (Eds.). (2014). ICT Innovations 2014: *World of Data (Vol. 311). Springer*.

Bonafonte, A., et al. (1996). Explicit segmentation of speech using Gaussian models. *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on, IEEE*.

Bourlard, H., et al. (1996). Copernicus and the ASR challenge-Waiting for Kepler. *Proceedings of the ARPA ASR Workshop Spring*, Citeseer.

Butzberger, J., et al. (1992). Spontaneous speech effects in large vocabulary speech recognition applications. *Proceedings of the workshop on Speech and Natural Language, Association for Computational Linguistics*.

Caballero-Morales, S.-O. (2013). Estimation of phoneme-specific HMM topologies for the automatic recognition of dysarthric speech. *Computational and mathematical methods in medicine* 2013.

Canvin, R. (2012, August 2012). *Speech and Language Therapy*. Retrieved 22 January, 2013, from http://www.bupa.com.mt/health-wellbeing/item/speech-and-language-therapy.

Caruntu, A., et al. (2005). Automatic silence/unvoiced/voiced classification of speech using a modified Teager energy feature. *Proceedings of the 2005 WSEAS international conference on Dynamical systems and control, World Scientific and Engineering Academy and Society (WSEAS).*

Casar, M. and J. A. Fonollosa (2008). *Overcoming HMM Time and Parameter Independence Assumptions for ASR*. 550.

Celce-Murcia, M. and L. McIntosh (1991). *Teaching English as a second or foreign language*, Newbury House.

Chelba, C., et al. (2012). *Large scale language modeling in automatic speech recognition*. arXiv preprint arXiv:1210.8440.

Christina, S. L., et al. (2012). *HMM-based speech recognition system for the dysarthric speech evaluation of articulatory subsystem*. Recent Trends In Information Technology (ICRTIT), 2012 International Conference on, IEEE.

Chou, F.-c., et al. (1998). *Automatic segmental and prosodic labeling of Mandarin speech database*. ICSLP.

Ciarcia, S. (1981). *Build your own Z80 computer: design guidelines and application notes*, Circuit Cellar.

Cleghorn, T. L. and N. M. Rugg (2011). *Comprehensive Articulatory Phonetics: A Tool for Mastering the World's Languages*, Daniel Ebuehi.

Clifford, H. C. and F. A. Swettenham (1894). *A Dictionary of the Malay Language*, authors at the Government's printing Office.

Clynes, A. and D. Deterding (2011). *Standard Malay (Brunei)*. Journal of the International Phonetic Association 41(02): 259-268.

Comerford, R., et al. (1997). The voice of the computer is heard in the land (and it listens too!)[speech recognition]. *Spectrum, IEEE* 34(12): 39-47.

Cosi, P., et al. (1991). *A preliminary statistical evaluation of manual and automatic segmentation discrepancies*. EuroSpeech.

Cox, S., et al. (1998). Techniques for accurate automatic annotation of speech waveforms. *ICSLP*.

Crestani, F. (2001). Towards the use of prosodic information for spoken document retrieval. P*roceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval, ACM*.

Das, B. K., et al. (2014). Detection of Voiced, Unvoiced and Silence Regions of Assamese Speech by Using Acoustic Features. *International Journal of Computer Trends and Technology (IJCTT)* Vol 14(No 2): pg 43-45.

Davis, K., et al. (1952). Automatic recognition of spoken digits. *The Journal of the Acoustical Society of America* 24(6): 637-642.

Davis, S. and P. Mermelstein (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *Acoustics, Speech and Signal Processing, IEEE Transactions* on 28(4): 357-366.

De Falco, I., et al. (2002). Mutation-based genetic algorithm: performance evaluation. *Applied Soft Computing* 1(4): 285-299.

DeMiller, A. L. (2000). *Linguistics: a guide to the reference literature,* Libraries Unlimited.

Denes, P. B. (1993). *The speech chain*, Macmillan.

Denver, I. (2012). *Speech disorders-children*.

Dermatas, E. S., et al. (1991). Fast endpoint detection algorithm for isolated word recognition in office environment. *Acoustics, Speech, and Signal Processing, 1991. ICASSP-91., 1991 International Conference on, IEEE*.

Deshmukh, O., et al. (2002). Acoustic-phonetic speech parameters for speaker-independent speech recognition. *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on, IEEE*.

Dewan, K. (2005). *Dewan Bahasa dan Pustaka*. Kuala Lumpur.

Diaconis, P. (1988). *Group representations in probability and statistics*. Lecture Notes-Monograph Series: i-192.

Díaz, F. C. and E. R. Banga (2006). A method for combining intonation modelling and speech unit selection in corpus-based speech synthesis systems. *Speech Communication* 48(8): 941-956.

Dodd, B. (2013). *Differential diagnosis and treatment of children with speech disorder*, John Wiley & Sons.

Dong, M., et al. (2009). *I2R text-to-speech system for Blizzard Challenge 2009*. Blizzard Challenge Workshop, Citeseer.

Dronkers, N. F. (1996). A new brain region for coordinating speech articulation. *Nature* 384(6605): 159-161.

Du, K. L. and N. S. Swamy (2013). *Neural Networks and Statistical Learning*, Springer London.

Dutoit, T. (1997). *An introduction to text-to-speech synthesis*, Springer Science & Business Media.

Edgington, M., et al. (1996). Overview of current text-to-speech techniques. II: Prosody and speech generation. *BT technology journal* 14(1): 84-99.

Eiben, A. E. and J. E. Smith (2003). *Introduction to evolutionary computing*, Springer Science & Business Media.

Eng, G. K. and A. M. bin Ahmad (2005). Applying Hybrid Neural Network For Malay Syllables Speech Recognition. *TENCON 2005 2005 IEEE Region 10, IEEE.*

Errede, S. (2014) *The Human Ear - Hearing*, Sound Intensity and Loudness Levels

Farago, A. and G. Lugosi (1989). An algorithm to find the global optimum of left-to-right hidden Markov model parameters. *Problems Of Control And Information Theory-Problemy Upravleniya I Teorii Informatsii* 18(6): 435-444.

Fey, M. E. (1986). *Language intervention with young children*, College-Hill Press San Diego, CA.

Fisher, W. M., et al. (1986). *The DARPA speech recognition research database: specifications and status*. Proc. DARPA Workshop on speech recognition.

Flanagan, J. L. (1972). *Speech analysis: Synthesis and perception*.

Fook, C., et al. (2012). Malay speech recognition in normal and noise condition. *Signal Processing and its Applications (CSPA), 2012 IEEE 8th International Colloquium on, IEEE.*

Foresman, B. R. (2008). *Acoustical Measurement of the Human Vocal Tract: Quantifying Speech & Throat-Singing*.

Forney Jr, G. D. (1973). The viterbi algorithm. *Procdngs of the IEEE* 61(3): 268-278.

Fosler-Lussier, E. (2003). *A tutorial on pronunciation modeling for large vocabulary speech recognition*. Text-and Speech-Triggered Information Access, Springer: 38-77.

Frankel, J. and S. King (2001). *ASR-articulatory speech recognition*.

Frederick Jelinek, R. L. M., and S. Roukos. (1992). Principles of lexical language modeling for speech recognition. *Adv. in Speech Signal Processing*: 651–699.

Fukada, T., et al. (1996). Speech recognition based on acoustically derived segment units. *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on, IEEE.*

Furui, S., et al. (2004). Speech-to-text and speech-to-speech summarization of spontaneous speech. *Speech and Audio Processing, IEEE Transactions* on 12(4): 401-408.

Gaikwad, S. K., et al. (2010). A review on speech recognition technique. *International Journal of Computer Applications* 10(3): 16-24.

Gales, M. and S. Young (2008). The application of hidden Markov models in speech recognition. *Foundations and Trends in Signal Processing* 1(3): 195-304.

Gambrell, L. B. and P. B. Jawitz (1993). Mental imagery, text illustrations, and children's story comprehension and recall. *Reading Research Quarterly:* 265-276.

Gargiulo, R. M. (2006). Special education in contemporary society : an introduction to exceptionality. *Belmont, CA*, Thomson/Wadsworth.

Garofolo, J. S., et al. (1993). DARPA TIMIT acoustic-phonetic continous speech corpus CD-ROM. *NIST speech disc 1-1.1. NASA* STI/Recon Technical Report N 93: 27403.

Gelfer, M. P. (1996). Survey of Communication Disorders: A Social and Behavioral Perspective, *McGraw-Hill*.

Gibson, J. W. and M. S. Hanna (1992). Introduction to human communication, McGraw-Hill *Humanities, Social Sciences & World Languages.*

Gil, D. (2007). A typology of stress, and where Malay/Indonesian fits in. *Eleventh International Symposium on Malay/Indonesian Linguistics*, Universitas Negeri Papua, Manokwari, Indonesia.

Glass, J. R. (1988). Finding Acoustic Regularities in Speech: Applications to Phonetic Recognition, *DTIC Document*.

Glykas, M. and P. Chytas (2004). Technology assisted speech and language therapy. *International journal of medical informatics* 73(6): 529-541.

Griffin, S., L. Wilson, E. Clark and S. McLeod (2000). *Speech pathology applications of Automatic Speech Recognition technology*, SST.

Gouvêa, E. (2014) *Training Acoustic Model For CMUSphinx.*

Greenberg, S. (1999). Speaking in shorthand–A syllable-centric perspective for understanding pronunciation variation. *Speech Communication* 29(2): 159-176.

Greenwood, M. and A. Kinghorn (1999). *Suving: Automatic silence/unvoiced/voiced classification of speech*. Undergraduate Coursework, Department of Computer Science, The University of Sheffield, UK.

Grigore, O., et al. (2011). Impaired Speech Evaluation using Mel-Cepstrum Analysis. *International Journal Of Circuits, Systems And Signal Processing*: 70-77.

Gruhn, R. E., et al. (2011). *Statistical Pronunciation Modeling for Non-Native Speech Processing*, Springer.

Gu, L. and S. A. Zahorian (2002). A new robust algorithm for isolated word endpoint detection. *Energy* 2: 1.

Gubrynowicz, R. and A. Wrzoskowicz (1993). Labeller-A System for Automatic Labelling of Speech Continuous Signal. *Third European Conference on Speech Communication and Technology.*

Gunawan, T. S., et al. (2011). *English digits speech recognition system based on hidden Markov Models.*

Gurjar, A. A. and S. A. Ladhake (2008). Time-frequency analysis of chanting sanskrit divine sound OM mantra. *Int. J. Comput. Sci. Network Secur* 8: 170-175.

Hain, T., S. Johnson, A. Tuerk, P. Woodland and S. Young (1998). *Segment generation and clustering in the HTK broadcast news transcription system*. Proc. 1998 DARPA Broadcast News Transcription and Understanding Workshop.

Hakkani-Tur, D., Riccardi, G., & Gorin, A. (2002). *Active learning for automatic speech recognition*. In Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on (Vol. 4, pp. IV-3904). IEEE.

Halevy, A., et al. (2009). The unreasonable effectiveness of data. *Intelligent Systems, IEEE* 24(2): 8-12.

Han, W., et al. (2006). An efficient MFCC extraction method in speech recognition. *Circuits and Systems, 2006. ISCAS 2006. Proceedings. 2006 IEEE International Symposium on, IEEE*.

Hansen, E. G. (2006). *Multilingual Phoneme Models for Rapid Speech Processing System Development*, DTIC Document.

Harris, D. (2012). *Google explains how more data means better speech recognition*. Retrieved 15 January, 2015, from https://gigaom.com/2012/10/31/google-explains-how-more-data-means-better-speech-recognition/.

Hassan, A. (1974). *The morphology of Malay*. Dewan Bahasa Dan Pustaka, Kementerian Pelajaran Malaysia (KPM).

Hayes, B. (2011). *Introductory phonology*, John Wiley & Sons.

Hirschberg, J. (2012). *Automatic Speech Recognition: An Overview.* Department of Computer Science, 1214 Amsterdam AvenueM/C 0401, 450 CS Building New York, NY 10027, Columbia University.

Hodson, B. W. (2007). *Evaluating and enhancing children's phonological systems*. Greenville, SC: Thinking Publications–University.

Hoffmann, S. and E. TIK (2009). *Automatic Phone Segmentation*. Corpora 3: 2.1.

Hon, H.-W., et al. (1989). Towards speech recognition without vocabulary-specific training. P*roceedings of the workshop on Speech and Natural Language*, Association for Computational Linguistics.

Honey, H. (2012, 4 February 2012). *Phoneme and Feature Theory.* Retrieved 15 February 2014, 2014, from http://www.slideshare.net/honeyravian1/phoneme-and-feature-theory.

Hoos, H. H. and T. Stützle (2004). *Stochastic local search: Foundations & applications*, Elsevier.

Huang, X. and M. Jack (1989). Unified techniques for vector quantization and hidden Markov modeling using semi-continuous models. *Acoustics, Speech, and Signal Processing, 1989. ICASSP-89., 1989 International Conference on, IEEE.*

Huang, X., et al. (2001). *Spoken language processing: A guide to theory, algorithm, and system development*, Prentice Hall PTR.

Huang, X. D. and M. A. Jack (1989). Semi-continuous hidden Markov models for speech signals. *Computer Speech & Language* 3(3): 239-251.

Huang, X. and L. Deng (2010). An overview of modern speech recognition. *Handbook of Natural Language Processing*: 339-366.

Hwang, W.-Y., et al. (2012). Effects of Speech-to-Text Recognition Application on Learning Performance in Synchronous Cyber Classrooms. *Educational Technology & Society* 15(1): 367-380.

Indurkhya, N. and F. J. Damerau (2010). *Handbook of Natural Language Processing*, Second Edition, Taylor & Francis.

Itakura, F. (1975). Minimum prediction residual principle applied to speech recognition. *Acoustics, Speech and Signal Processing, IEEE Transactions* on 23(1): 67-72.

Ittycheriah, A., et al. (2001). *Apparatus and methods for identifying homophones among words in a speech recognition system*, Google Patents.

J. Stephen, H. A. (2014). Speech Recognition Using Genetic Algorithm. *International Journal of Computer Engineering & Technology* (IJCET) Vol 5.(Issue 5): pp. 76-81.

Jabloun, F., et al. (1999). Teager energy based feature parameters for speech recognition in car noise. *Signal Processing Letters, IEEE* 6(10): 259-261.

Jadhav, S., et al. V*oice Activated Calculator.*

Jamilah, P. (2014). *Sekolah Kebangsaan Pendidikan Khas (SKPK) Interview*. SKPK Johor Bahru.

Jang, J.-S. R. (2009). *Audio signal processing and recognition*. Information on http://www. cs. nthu. edu. tw/~ jang.

Jang, P. J. and A. G. Hauptmann (1999). Learning to recognize speech by watching television. *Intelligent Systems and their Applications, IEEE* 14(5): 51-58.

Jarifi, S., et al. (2008). A fusion approach for automatic speech segmentation of large corpora with application to speech synthesis. *Speech Communication* 50(1): 67-80.

Jelinek, F. (1997). *Statistical methods for speech recognition*, MIT press.

Jelinek, F., et al. (1975). Design of a linguistic statistical decoder for the recognition of continuous speech. *Information Theory, IEEE Transactions* on 21(3): 250-256.

John, H. (1992). Holland, *Adaptation in natural and artificial systems,* MIT Press, Cambridge, MA.

John-Paul Hosom, R. C., Mark Fanty, Johan Schalkwyk, Yonghong Yan, Wei Wei (1998) *Training Neural Networks for Speech Recognition*. Second Draft,

Johnson, N. L. and N. Balakrishnan (1998). Advances in the Theory and Practice of Statistics. *Technometrics* 40(2): 165-165.

Joutsenvirta, A. (2009). *Hertz, Cent and Decibel*. Retrieved 3 December, 2014, from http://www2.siba.fi/akustiikka/?id=38&la=en.

Jurafsky, D. (2014). *Lecture 3: ASR: HMMs, Forward, Viterbi* Stanford University, CS 224s/LINGUIST 285: Spoken Language Processing.

Jurafsky, D. and J. H. Martin (2000). *Speech & language processing,* Pearson Education India.

Kamarudin, N., et al. (2014). Feature extraction using Spectral Centroid and Mel Frequency Cepstral Coefficient for Quranic Accent Automatic Identification. *Research and Development (SCOReD)*, 2014 IEEE Student Conference.

Kao, Y. H. (2002). *N-best search for continuous speech recognition using viterbi pruning for non-output differentiation states*, Google Patents.

Kemsley, S. (2008). *Word List: Homonyms.* abctools®. Union Lake, MI

Kesarkar, M. P. (2003). *Feature extraction for speech recognition*. Tech. Credit Seminar Report, Electronic Systems Group, EE. Dept, IIT Bombay.

Khaled, F. and S. Fawzy (2013). Comparative Evaluation of Different Mel-Frequency Cepstral Coefficient Features on Qura'nic-Arabic Verification. *International Journal of Computer Applications* 84(16): 44-48.

KKM (2014). *Wordlist for Speech Theraphy*. K. K. Malaysia.

Klatt, D. H. (1977). Review of the ARPA speech understanding project. *The Journal of the Acoustical Society of America* 62(6): 1345-1366.

Klatt, D. H. (1980). *Overview of the ARPA speech understanding project*. Trends in Speech Recognition', Prentice-Hall, Englewood Cliffs, NJ: 249-271.

Kleynhans, N. T., Barnard, E. (2015). *Efficient data selection for ASR*. Language Resources and Evaluation, 49(2): 327-353.

KM, R. K. and S. Ganesan (2011). Comparison of multidimensional MFCC feature vectors for objective assessment of stuttered disfluencies. *age* 2(05): 854-860.

Koffka, K. (2013). *Principles of Gestalt psychology*, Routledge.

Kottegoda, N. T. and R. Rosso (2009). *Applied Statistics for Civil and Environmental Engineers*, Wiley.

Kotz, S. and N. L. Johnson (1985). *Multiple range and associated test procedures Encyclopedia of Statistical Sciences 5*, Wiley, New York.

KPM (2002). *Maklumat Pendidikan Khas Kuala Lumpur, Jabatan Pendidikan Khas Kementerian Pendidikan Malaysia*, Kementerian Pendidikan Malaysia.

Kumar, C. S. and P. M. Rao (2011). Design of an automatic speaker recognition system using MFCC, Vector quantization and LBG algorithm. *Ch. Srinivasa Kumar et al./International Journal on Computer Science and Engineering (IJCSE)* 3(8).

Kuo, J.-W. and H.-M. Wang (2006). A minimum boundary error framework for automatic phonetic segmentation. *Chinese Spoken Language Processing*, Springer: 399-409.

Kurzweil, R. (2012). *How to Create a Mind: The Secret of Human Thought Revealed*, Penguin Group US.

Kwong, S., et al. (2001). Optimisation of HMM topology and its model parameters by genetic algorithms. *Pattern recognition* 34(2): 509-522.

Ladefoged, P. and K. Johnson (2014). *A course in phonetics*, Cengage learning.

Lamberts, K. and R. Goldstone (2004). *Handbook of cognition*, Sage.

Laver, J. (1994). *Principles of phonetics*, Cambridge University Press.

Lee, G. G., Kim, H. K., Jeong, M., & Kim, J. H. (Eds.). (2015). *Natural Language Dialog Systems and Intelligent Assistants*. Springer.

Lee, C.-y. and J. Glass (2012). A nonparametric Bayesian approach to acoustic model discovery. *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*: Long Papers-Volume 1, Association for Computational Linguistics.

Lee, I. (2011). *The application of speech recognition technology for remediating the writing difficulties of students with learning disabilities*, University of Washington.

Lee, K.-F. (1989). *Automatic Speech Recognition: The Development of the Sphinx Recognition System*, Springer.

Lee, K.-F., et al. (1990). An overview of the SPHINX speech recognition system. *Acoustics, Speech and Signal Processing, IEEE Transactions* on 38(1): 35-45.

Levin, E. (1991). *Modeling time varying systems using hidden control neural architecture*. Advances in neural information processing systems.

Li, J. X. L. Z. C. (2007). Optimization of hidden Markov model by a genetic algorithm for web information extraction. *ISKE-2007 Proceedings.*

Liboh, H. (2011). *Computer-based malay speech analysis system for speech theraphy system*, Universiti Teknologi Malaysia, Faculty of Electrical Engineering. Master. Thesis

Liu, W. and H. Weisheng (2010). Improved Viterbi algorithm in continuous speech recognition. *Computer Application and System Modeling (ICCASM), 2010 International Conference on, IEEE*.

Livescu, K., et al. (2012). Subword modeling for automatic speech recognition: Past, present, and emerging approaches. *Signal Processing Magazine, IEEE* 29(6): 44-57.

Ljolje, A., et al. (1997). *Automatic speech segmentation for concatenative inventory selection. Progress in speech synthesis*, Springer: 305-311.

Ljolje, A. and M. Riley (1991). Automatic segmentation and labeling of speech. *Acoustics, Speech, and Signal Processing, 1991. ICASSP-91., 1991 International Conference on, IEEE.*

Lo, H.-Y. and H.-M. Wang (2007). Phonetic boundary refinement using support vector machine. *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on, IEEE*.

Lowerre, B. (1990). *The HARPY speech understanding system*. Readings in speech recognition, Morgan Kaufmann Publishers Inc.

Lowerre, B. T. (1976). *The HARPY speech recognition system.*

Malfrère, F., et al. (1998). *Phonetic alignment: speech synthesis based vs. hybrid HMM/ANN*. ICSLP.

Malfrere, F. and T. Dutoit (1997). High-quality speech synthesis for phonetic speech segmentation. *Fifth European Conference on Speech Communication and Technology*.

Maris, Y. (1966). *The Malay sound system*, University of Malaya.

Martirosian, O. and M. Davel (2007). Error analysis of a public domain pronunciation dictionary, *18th Annual Symposium of the Pattern Recognition Association of South Africa (PRASA 2007)*.

Marks, J. A. (1988). Real time speech classification and pitch detection. *Communications and Signal Processing, 1988. Proceedings., COMSIG 88. Southern African Conference on, IEEE.*

Matoušek, J. (2009). *Automatic pitch-synchronous phonetic segmentation with context-independent HMMs*. Text, Speech and Dialogue, Springer.

Matoušek, J. and J. Romportl (2008). *Automatic pitch-synchronous phonetic segmentation*.

Matsui, T. and S. Furui (1993). Concatenated phoneme models for text-variable speaker recognition. *Acoustics, Speech, and Signal Processing, 1993. ICASSP-93., 1993 IEEE International Conference on, IEEE.*

Mazenan, M. N. and T.-S. Tan (2013). Malay Alveolar Vocabulary Design for Malay Speech Therapy System. *Recent Advances in Electrical Engineering Series*(11).

Mazenan, M. N. and T.-S. Tan (2014). *Malay Wordlist Modeling for Articulation Disorder Patient by Using Computerized Speech Diagnosis System*.

McKeown, K. (2010). *N-Grams and Corpus Linguistics.* Retrieved October, 2014, from http://www.cs.columbia.edu/~kathy/NLP/ClassSlides/Class3-ngrams09/ngrams.pdf.

Meir D., Z. Y. (2007). *Saya Speech Recognition. Mini Project*. Department of Computer Science, Ben-Gurion University, Ben-Gurion University.

Michailovsky, B. (1994). Manner vs place of articulation in the Kiranti initial stops. *Current issues in Sino-Tibetan linguistics*.

Miesenberger, K., et al. (2014). Computers Helping People with Special Needs: *14th International Conference, ICCHP 2014, Paris, France,* July 9-11, 2014 : Proceedings, Springer.

Misman, S. Z. (2014). *Interview on Speech Disorder Analysis and Statistic for HSA*. M. N. Mazenan.

Mitchell, C. D., et al. (1995). Using explicit segmentation to improve HMM phone recognition. Acoustics, Speech, and Signal Processing, *IEEE International Conference on, IEEE.*

Morrison, E. M. (2014). *Method and system for teaching and practicing articulation of targeted phonemes*, Google Patents.

Mosleh, M., et al. (2009). *A synergy between HMM-GA based on stochastic cellular automata to accelerate speech recognition*. IEICE Electronics Express 6(18): 1304-1311.

Mporas, I., et al. (2010). Speech segmentation using regression fusion of boundary predictions. *Computer Speech & Language* 24(2): 273-288.

Muda, L., et al. (2010). *Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques*. arXiv preprint arXiv:1003.4083.

Muthusamy, H., et al. (2012). *Comparison of LPCC and MFCC for isolated Malay speech recognition*.

Nagroski, A., et al. (2003). In search of optimal data selection for training of automatic speech recognition systems. A*utomatic Speech Recognition and Understanding, 2003. ASRU'03. 2003 IEEE Workshop on, IEEE.*

Navarro-Newball, A. A., et al. *Talking to Teo: Video game supported speech therapy*. Entertainment Computing 5.4 (2014): 401-412.

Neilsen, M. (2009). Supporting struggling writers with the use of voice recognition software in class. *Literacy* 49(1): 30-39.

Nelson, K. E. (1989). *Strategies for first language teaching*.

Ney, H. (1984). The use of a one-stage dynamic programming algorithm for connected word recognition. *Acoustics, Speech and Signal Processing, IEEE Transactions* on 32(2): 263-271.

Ng, V. (2008). *Statistical natural language processing*.

Nisbet, P., et al. (2002). *Introducing Speech Recognition in Schools: Using Dragon NaturallySpeaking*, CALL Centre.

(NIST), N. I. o. s. a. t. (2008). *NIST Scoring for ASR*. USA, U.S. Department of Commerce.

Nitin N Lokhande, N. S. N., Pratap S Vikhe (2011). Voice Activity Detection Algorithm for Speech Recognition Applications. *International Conference in Computational Intelligence (ICCIA) IJCA*: pg 5-7.

Nizam Mazenan, T. T. (2014). Variation Word Test and Lexical Unit Selection in Malay Corpus Design for Articulation Disorder Screening with Computerized System. *3rd International Conference on Applied and Computational Mathematics (ICACM '14). Geneva, Switzerland, WSEAS* - World Scientific and Engineering Academy and Society.

Nong, T. H., et al. (2001). Classification of Malay speech sounds based on place of articulation and voicing using neural networks. *TENCON 2001. Proceedings of IEEE Region 10 International Conference on Electrical and Electronic Technology, IEEE.*

Ogbureke, K. U. and J. Carson-Berndsen (2009). Improving initial boundary estimation for HMM-based automatic phonetic segmentation. *INTERSPEECH, Citeseer.*

O'Beirne, R. (2006). Berkshire Encyclopedia of Human-Computer Interaction. R*eference Reviews* 20(5): 15-16.

O'shaughnessy, D. (1987). *Speech communication: human and machine*, Universities press.

Okumura, A. (2014). *What is Speech Recognition Technology?*. Retrieved January 2014, 2014, from http://www.nec.com/en/global/rd/innovative/speech/01.html.

Olson, H. F. (1967). *Music, physics and engineering*, Courier Dover Publications.

Omar, A. H. (1983). *The Malay peoples of Malaysia and their language*s, Dewan Bahasa dan Pustaka, Kementerian Pelajaran Malaysia.

Omar, A. H. (2008). *Ensiklopedia Bahasa Melayu*, Dewan Bahasa dan Pustaka.

Omura, J. (2006). On the Viterbi decoding algorithm. *IEEE Trans. Inf. Theor.* 15(1): 177-179.

Openshaw, J., et al. (1993). A comparison of composite features under degraded speech in speaker recognition. *Acoustics, Speech, and Signal Processing, 1993. ICASSP-93., 1993 IEEE International Conference on, IEEE.*

Oshodi, B. (2013). *A cross-language study of the speech sounds in Yorùbá and Malay: Implications for Second Language Acquisition*. Issues In Language Studies: 1.

Otto, B. (2014). *Language Development in Early Childhood Education*, Pearson.

Parente, R., et al. (2004). *An analysis of the implementation and impact of speech-recognition technology in the healthcare sector.* Perspectives in health information management/AHIMA, American Health Information Management Association 1.

Paul D. Nisbet, A. W., Dr. Stuart Aitken (2005). *Speech Recognition for Students with Disabilities*. Inclusive and Supportive Education Congress ISEC 2005. University of Strathclyde, Glasgow, Scotland, Inclusive Technology Ltd.

Paul, D. B. (1992). An efficient A* stack decoder algorithm for continuous speech recognition with a stochastic language model. *Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on, IEEE.*

Paul, D. B. (1990). Speech recognition using hidden markov models. *The Lincoln Laboratory Journal* 3(1): 41-62.

Pedretti, L. W. and M. B. Early (2001). *Occupational Therapy: Practice Skills for Physical Dysfunction*, Mosby.

Peinado, A., et al. (1994). *Use of multiple vector quantisation for semicontinuous-HMM speech recognition*. Vision, Image and Signal Processing, IEE Proceedings-, IET.

Pellom, B. and K. Hacioglu (2001). *Sonic: The university of colorado continuous speech recognizer*. University of Colorado, tech report# TR-CSLR-2001-01, Boulder, Colorado.

Pellom, B. L. and J. H. Hansen (1998). *Automatic segmentation of speech recorded in unknown noisy channel characteristics*. Speech Communication 25(1): 97-116.

Phoon, H. S., et al. (2014). *Consonant acquisition in the Malay language: A cross-sectional study of preschool aged malay children*. Clinical linguistics & phonetics 28(5): 329-345.

Pieraccini, R. (2012). *From AUDREY to Siri. Is speech recognition a solved problem?* The Annual AVIOS Conference. Berkeley The Internatonal Computer Science Institute.

Pieraccini, R. (2012). *The voice in the machine: building computers that understand speech*, MIT Press.

Pieraccini, R., et al. (1992). A speech understanding system based on statistical representation of semantics. *Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on, IEEE.*

Pieraecini, R. and E. Levin (1995). *A learning approach to natural language understanding. Speech Recognition and Coding*, Springer: 139-156.

Pitsikalis, V. and P. Maragos (2009). Analysis and classification of speech signals by generalized fractal dimension features. *Speech Communication* 51(12): 1206-1223.

Poritz, A. and A. Richter (1986). On hidden Markov models in isolated word recognition. *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'86., IEEE.*

Powers, M. H. (1971). *Clinical and educational procedures in functional disorders of articulation*. Handbook of Speech Pathology and Audiology. New York: Appleton-Century-Crofts.

Pring, T. (2005). *Research methods in communication disorders*, Wiley.

Priyam, R., et al. (2013). Artificial Intelligence Applications for Speech Recognition. *Proceedings of the Conference on Advances in Communication and Control Systems-2013, Atlantis Press.*

Rabiner, L. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* 77(2): 257-286.

Rabiner, L. and B.-H. Juang (1986). An introduction to hidden Markov models. *ASSP Magazine, IEEE* 3(1): 4-16.

Rabiner, L., et al. (1986). A model-based connected-digit recognition system using either hidden Markov models or templates. *Computer Speech & Language* 1(2): 167-197.

Rabiner, L. R. and B.-H. Juang (1993). *Fundamentals of speech recognition*, PTR Prentice Hall Englewood Cliffs.

Rabiner, L. R. and B.-H. Juang (2006). Speech recognition: Statistical methods. *Encyclopedia of language & linguistics*: 1-18.

Rabiner, L. R. and M. R. Sambur (1975). An algorithm for determining the endpoints of isolated utterances. *Bell System Technical Journal* 54(2): 297-315.

Rabiner, L. R. and R. W. Schafer (1978). *Digital processing of speech signals*, Prentice-hall Englewood Cliffs.

Raflesia, S. (2008). *Learning Support and Intervention Services.* Retrieved 8 April 2013, from www.srirafelsia.com.

Raghavan, P. (1986). Probabilistic construction of deterministic algorithms: approximating packing integer programs. *Foundations of Computer Science, 1986., 27th Annual Symposium on, IEEE.*

Rahman, F. D., et al. (2014). *Automatic speech recognition system for Malay speaking children*. Student Project Conference (ICT-ISPC), 2014 Third ICT International, IEEE.

Raminah, S. and S. Rahim (1987). *Kajian Bahasa untuk Pelatih Maktab Perguruan*, Petaling Jaya: Penerbit Fajar Bakti Sdn. Bhd.

Ranchod, E. and N. J. Mamede (2002). *Advances in Natural Language Processing:* Third International Conference, PorTAL 2002, Faro, Portugal, June 23-26, 2002. Proceedings, Springer Science & Business Media.

Razak, Z., et al. (2008). *Quranic verse recitation feature extraction using mel-frequency cepstral coefficient (MFCC)*. 4th International Colloquium on Signal Processing and its Applications (CSPA 2008).

Rekha, J. U., Chatrapati, K. S., & Babu, A. V. (2015). *A Fast Algorithm for HMM Training using Game Theory for Phoneme Recognition.* International Journal of Computer Applications, 118(11).

Reynolds, D. A. and R. C. Rose (1995). Robust text-independent speaker identification using Gaussian mixture speaker models. *Speech and Audio Processing, IEEE Transactions* on 3(1): 72-83.

Riedhammer, K., et al. (2012). Revisiting semi-continuous hidden Markov models. *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on, IEEE.*

Riley, M. D. and A. Ljolje (1996). Automatic generation of detailed pronunciation lexicons. Automatic Speech and Speaker Recognition, *Springer:* 285-301.

Riley, M., et al. (1999). Stochastic pronunciation modelling from hand-labelled phonetic corpora. *Speech Communication* 29(2): 209-224.

Rodríguez-Andina, J., et al. (2001). A FPGA-based Viterbi algorithm implementation for speech recognition systems. *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on, IEEE.*

Roebuck, K. (2012). *Predictive Analysis: High-impact Emerging Technology - What You Need to Know:* Definitions, Adoptions, Impact, Benefits, Maturity, Vendors, Emereo Publishing.

Ronnel Ivan A. Casil, E. D. D., Maria Elena C. Manamparan, Bahareh Ghorban Nia, and Angelo A. Beltran, Jr. (2014). A DSP Based Vector Quantized Mel Frequency Cepstrum Coefficients for Speech Recognition. *Institute of Electronics Engineers of the Philippines (IECEP) Journal* Vol. 3(No. 1): pp. 1-5.

Rosdi, F. and R. N. Ainon (2008). Isolated malay speech recognition using Hidden Markov Models. *Computer and Communication Engineering, 2008. ICCCE 2008. International Conference on, IEEE.*

Rosenfield, R. (2000). *Two decades of statistical language modeling: Where do we go from here?.*

Rubin, Zachary, Sri Kurniawan, and Travis Tollefson. *Results from Using Automatic Speech Recognition in Cleft Speech Therapy with Children*. Computers Helping People with Special Needs. Springer International Publishing, 2014. 283-286.

Rybach, D., C. Gollan, R. Schluter and H. Ney (2009). Audio segmentation for speech recognition using segment features. *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on, IEEE.*

Sakoe, H. and S. Chiba (1970). *A similarity evaluation of speech patterns by dynamic programming.* Nat. Meeting of Institute of Electronic Communications Engineers of Japan.

Salaja, R. T., Flynn, R., Russell, M. (2014). *Automatic speech recognition using artificial life*.

Samsudin, N. (2007). *Study on Phonetic Context of Malay Syllables Towards the Development of Malay Speech Synthesizer*. Universiti Sains Malaysia (USM). Pulau Pinang, Universiti Sains Malaysia (USM). Master of Science. Thesis

Sánchez-Martínez, F., et al. (2006). Speeding up target-language driven part-of-speech tagger training for machine translation. *MICAI 2006: Advances in Artificial Intelligence*, Springer: 844-854.

Sandanalakshmi, R., et al. (2013). *Speaker Independent Continuous Speech to Text Converter for Mobile Application.* arXiv preprint arXiv:1307.5736.

Sariyan, A. and S. Karim (1984). *Isu-isu bahasa Malaysia,* Penerbit Fajar Bakti.

Saz, O., S.-C. Yin, E. Lleida, R. Rose, C. Vaquero and W. R. Rodríguez (2009). Tools and technologies for computer-aided speech and language therapy. *Speech Communication* 51(10): 948-967.

Scarrott (2014). *The Fifth Generation Computer Project: State of the Art Report 11:1*, Elsevier Science.

Scotland, C. (2014). *Speech Recognition in Schools Project*. Retrieved 15 October, 2014, from http://www.callscotland.org.uk/What-we-do/Projects/Speech-Recognition/.

Sebastiaan Hoekstra, G. v. d. H., Jeroen Rijnbout, Romy Visser (2013). *TTTV2013project: SIRI*. 15 February 2015, from https://sites.google.com/site/tttv2013project/home/teammembers.

Secord, W., S. Boyce, J. Donohue, R. Fox and R. Shine (2007). *Eliciting sounds: Techniques and strategies for clinicians*, Cengage Learning.

Seigel, M. S. (2013). *Confidence Estimation for Automatic Speech Recognition Hypotheses*.

Seman, N., et al. (2010). An evaluation of endpoint detection measures for malay speech recognition of an isolated words. *Information Technology (ITSim), 2010 International Symposium in, IEEE.*

Seman, N. and K. Jusoff (2008). Acoustic pronunciation variations modeling for standard Malay speech recognition. *Computer and Information Science* 1(4): p112.

Shadiev, R., et al. (2014). Review of Speech-to-Text Recognition Technology for Enhancing Learning. *Journal of Educational Technology & Society* 17(4).

Shannon, C. E. (1948). Bell System Tech. J. 27 (1948) 379; CE Shannon. *Bell System Tech. J* 27: 623.

Shanthi Therese, S. and C. Lingam (2013). *Review of Feature Extraction Techniques in Automatic Speech Recognition*.

Shinoda, K. and T. Watanabe (2000). 1) MDL-based context-dependent subword modeling for speech recognition. *The Journal of the Acoustical Society of Japan (E)* 21(2): 79-86.

Siegler, M. (2011) *Siri, Do You Use Nuance Technology? Siri: I'm Sorry, I Can't Answer That.*

Sinclair, M., et al. (2014). A semi-Markov model for speech segmentation with an utterance-break prior. *Fifteenth Annual Conference of the International Speech Communication Association.*

Slobada, T. and A. Waibel (1996). Dictionary learning for spontaneous speech recognition. *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on, IEEE.*

Soegaard, M. (2010). Gestalt principles of form perception. *Interaction-Design. org*: 8

Spall, J. C. (2005). *Introduction to stochastic search and optimization: estimation, simulation, and control*, John Wiley & Sons.

Sparreboom, Marloes, et al. *Long-term outcomes on spatial hearing, speech recognition and receptive vocabulary after sequential bilateral cochlear implantation in children.* Research in dev. disabilities 36 (2015): 328-337.

Stahlberg, F., et al. (2014). *Towards Automatic Speech Recognition Without Pronunciation Dictionary, Transcribed Speech and Text Resources in the Target Language Using Cross-Lingual Word-to-Phoneme Alignment*. Spoken Language Technologies for Under-Resourced Languages.

Stemmer, G., et al. (2001). Multiple time resolutions for derivatives of Mel-frequency cepstral coefficients. *Automatic Speech Recognition and Understanding, 2001. ASRU'01. IEEE Workshop on, IEEE.*

Stevens, S. S., et al. (1937). A scale for the measurement of the psychological magnitude pitch. *The Journal of the Acoustical Society of America* 8(3): 185-190.

Stolcke, A., et al. (2006). Recent innovations in speech-to-text transcription at SRI-ICSI-UW. *Audio, Speech, and Language Processing, IEEE Transactions* on 14(5): 1729-1744.

Ström, N. (1997). *Automatic continuous speech recognition with rapid speaker adaptation for human/machine interaction*. Unpublished doctoral dissertation, KTH.

Suuny, S., S. David Peter and K. P. Jacob (2013). Performance of Different Classifiers in Speech Recognition. *International Journal of Research in Engineering and Technology* 2(4): 7.

Swee, T. T. and S. H. S. Salleh (2008). Corpus-based Malay text-to-speech synthesis system. *Communications, 2008. APCC 2008. 14th Asia-Pacific Conference on, IEEE.*

Swee, T. T., et al. (2010). Speech pitch detection using short-time energy. *Computer and Communication Engineering (ICCCE), 2010 International Conference on, IEEE.*

Tabassum, I. (2010). *Artificial Intelligence for Speech Recognition.* Department of Electronics and Communication, HKBK College of Engineering.

Takami, J.-i. and S. Sagayama (1992). A successive state splitting algorithm for efficient allophone modeling. *Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on, IEEE.*

Tan, T.-S., A. Ariff, C.-M. Ting and S.-H. Salleh (2007). Application of Malay speech technology in Malay speech therapy assistance tools. *Intelligent and Advanced Systems, 2007. ICIAS 2007. International Conference on, IEEE.*

Takara, T., et al. (1997). *Isolated word recognition using the HMM structure selected by the genetic algorithm*. Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on, IEEE.

Tanaka, H., et al. (2014). Linguistic and Acoustic Features for Automatic Identification of Autism Spectrum Disorders in Children's Narrative. *ACL 2014: 88*.

Taylor, P. A. (1994). *A phonetic model of English intonation*, Citeseer.

Teoh, B. S. and D. B. dan Pustaka (1994). *The sound system of Malay revisited*, Dewan Bahasa dan Pustaka, Ministry of Education, Malaysia.

Thomas, P. J. and F. F. Carmack (1990). *Speech and Language: Detecting and Correcting Special Needs*, Allyn and Bacon.

Ting, C.-M., et al. (2007). Automatic phonetic segmentation of Malay speech database. *Information, Communications & Signal Processing, 2007 6th International Conference on, IEEE.*

Ting, H., J. Yunus, S. Vandort and L. Wong (2003). Computer-based Malay articulation training for Malay plosives at isolated, syllable and word level. *Information, Communications and Signal Processing, 2003 and Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint Conference of the Fourth International Conference on, IEEE.*

Ting, H. N., et al. (2001). Malay syllable recognition based on multilayer perceptron and dynamic time warping. *Signal Processing and its Applications, Sixth International, Symposium on. 2001, IEEE.*

Ting, H. N. and J. Yunus (2004). Speaker-independent Malay vowel recognition of children using multi-layer perceptron. *TENCON 2004. 2004 IEEE Region 10 Conference, IEEE.*

Ting, H. N., et al. (2002). Speaker-independent phonation recognition for Malay Plosives using neural networks. *International joint conference on neural networks.*

Ting, H. N., et al. (2002). Speaker-independent Malay isolated sounds recognition. *Neural Information Processing, 2002. ICONIP'02. Proceedings of the 9th International Conference on, IEEE.*

Tischer, S. (2009). *Methods, Systems, and Products for Synthesizing Speech*, Google Patents.

Toledano, D. T., et al. (2003). Automatic phonetic segmentation. *Speech and Audio Processing, IEEE Transactions* on 11(6): 617-625.

Tuzun, O., et al. (1994). Comparison of parametric and non-parametric representations of speech for recognition. *Electrotechnical Conference, 1994. Proceedings., 7th Mediterranean, IEEE.*

Trevino-Zimmerman, K. (2008). *The Impact of Early Intervention Programs on Young Children with Speech and Language Delay: Comparison and Analysis*.

van Hemert, J. P. (1991). Automatic segmentation of speech. *Signal Processing, IEEE Transactions* on 39(4): 1008-1012.

Van Riper, C. (1978). *Speech correction: Principles and methods*.

Van Riper, C. G. and J. V. Irwin (1958). *Voice and articulation*, Pitman Medical Publishing Company.

Van Riper, C. and R. L. Erickson (1996). *Speech correction: An introduction to speech pathology and audiology*, Allyn and Bacon.

Varadarajan, B., et al. (2008). *Unsupervised learning of acoustic sub-word units*. Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers, Association for Computational Linguistics.

Varga, A. P. and R. Moore (1990). Hidden Markov model decomposition of speech and noise. *Acoustics, Speech, and Signal Processing, 1990. ICASSP-90., 1990 International Conference on, IEEE.*

Varile, G. B., et al. (1997). *Survey of the state of the art in human language technology*, Cambridge University Press Cambridge.

Veeravalli, A., et al. (2005). A tutorial on using hidden markov models for phoneme recognition. *System Theory, 2005. SSST'05. Proceedings of the Thirty-Seventh Southeastern Symposium on, IEEE.*

Vergin, R., et al. (1996). Compensated mel frequency cepstrum coefficients. *Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on, IEEE.*

Ververidis, D. and C. Kotropoulos (2006). Emotional speech recognition: Resources, features, and methods. *Speech Communication* 48(9): 1162-1181.

Vintsyuk, T. (1968). Speech discrimination by dynamic programming. *Cybernetics and Systems Analysis* 4(1): 52-57.

Viterbi, A. J. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *Information Theory, IEEE Transactions* on 13(2): 260-269.

Waibel, A. and K.-F. Lee (1990). *Readings in Speech Recognition*, Morgan Kaufmann Publishers.

Wald, M. and K. Bain (2008). Universal access to communication and learning: the role of automatic speech recognition. *Universal Access in the Information Society* 6(4): 435-447.

Wang, J., et al. (2013). *Word recognition from continuous articulatory movement time-series data using symbolic representations*.

Wang, J.-C., et al. (2002). Chip design of MFCC extraction for speech recognition. *INTEGRATION, the VLSI journal* 32(1): 111-131.

Wang, J., Samal, A., Green, J. R., & Rudzicz, F. (2012). *Whole-word recognition from articulatory movements for silent speech interfaces.*

Warren, S. F. and P. J. Yoder (1996). Enhancing communication and language development in young children with developmental delays and disorders. *Peabody Journal of Education* 71(4): 118-132.

Welch, L. R. (2003). Hidden Markov models and the Baum-Welch algorithm. *IEEE Information Theory Society Newsletter* 53(4): 10-13.

Werker, J. F. (1995). Exploring developmental changes in cross-language speech perception. *Language: an invitation to cognitive science* 1: 87-106.

Wester, M. and E. Fosler-Lussier (2000). *A comparison of data-derived and knowledge-based modeling of pronunciation variation*.

Wickramarachi, P. (2003). Effects of Windowing on the Spectral Content of a Signal. *Sound and Vibration* 37(1): 10-13.

Wightman, C. W. and D. T. Talkin (1997). The aligner: Text-to-speech alignment using Markov models. *Progress in speech synthesis, Springer*: 313-323.

Wolf, J. J. and W. Woods (1977). The HWIM speech understanding system. *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'77., IEEE.*

Wolraich, M. (2003). *Disorders of development and learning*, PMPH-USA.

Woo, P. and H. Teoh (2007). An investigation of cognitive and behavioural problems in children with attention deficit hyperactive disorder and speech delay. *Malaysian Journal of Psychiatry* 16(2): 50-58.

Woodland, T. H. P. (1998). *Segmentation and classification of broadcast news audio*.

Woods, W. A. (1983). Language processing for speech understanding. *NASA STI/Recon Technical Report N* 84: 10423.

Wu, S., et al. (2011). Automatic speech emotion recognition using modulation spectral features. *Speech Communication* 53(5): 768-785.

Wu, Y., et al. (2007). Data selection for speech recognition. *Automatic Speech Recognition & Understanding, 2007. ASRU. IEEE Workshop on, IEEE.*

Yakcoub, M. S., et al. (2008). Speech assistive technology to improve the interaction of dysarthric speakers with machines. *Communications, Control and Signal Processing, 2008. ISCCSP 2008. 3rd International Symposium on, IEEE.*

Yamada, T. and S. Sagayama (1996). LR-parser-driven Viterbi search with hypotheses merging mechanism using context-dependent phone models. *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on, IEEE.*

Yamamoto, K., et al. (2006). Robust endpoint detection for speech recognition based on discriminative feature extraction. *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on, IEEE.*

Yee, C. S. and A. M. Ahmad (2008). Malay language text-independent speaker verification using NN-MLP classifier with MFCC. *Electronic Design, 2008. ICED 2008. International Conference on, IEEE.*

Yehia, H. and F. Itakura (1996). A method to combine acoustic and morphological constraints in the speech production inverse problem. *Speech Communication* 18(2): 151-17

Ying, G., et al. (1993). Endpoint detection of isolated utterances based on a modified Teager energy measurement. *Acoustics, Speech, and Signal Processing, 1993. ICASSP-93., 1993 IEEE International Conference on, IEEE.*

Yoder, B. W. (2002). *Spontaneous speech recognition using HMM*, Massachusetts Institute of Technology.

Yong, L. C. and T. T. Swee (2014). Low footprint high intelligibility Malay speech synthesizer based on statistical data. *Journal of Computer Science* 10(2): 316-324.

Young, S. (1996). A review of large-vocabulary continuous-speech. *Signal Processing Magazine, IEEE* 13(5): 45.

Young, S., et al. (2006). *The HTK Book (for HTK Version. 3.4), Cambridge University Engineering Department (2006)*. Resume (April 6, 2009) Dipl.-Inform.(MS) Wolfgang Macherey.

Yu-jing, C. (2008). *Strategies for improving efficiency of Speech Recognition*, Beijing University of Posts and Telecommunications. Master. Thesis

Zhang, D., et al. (1997). A robust and fast endpoint detection algorithm for isolated word recognition. *Intelligent Processing Systems, 1997. ICIPS'97. 1997 IEEE International Conference on, IEEE.*

Zhili, L., et al. (2010). A study and application of speech recognition technology in primary and secondary school for deaf/hard of hearing students. *Proceedings of the 4th International Convention on Rehabilitation Engineering & Assistive Technology, Singapore Therapeutic, Assistive & Rehabilitative Technologies (START) Centre.*

Ziegler, J. C., et al. (2001). Identical words are read differently in different languages. *Psychological Science* 12(5): 379-384.

Ziółko, B., S. Manandhar, R. C. Wilson and M. Ziółko (2006). Wavelet method of speech segmentation. *Proceedings of 14th European Signal Processing Conference EUSIPCO, Florence.*

Zissman, M. A. (1996). Comparison of four approaches to automatic language identification of telephone speech. *IEEE Transactions on Speech and Audio Processing* 4(1): 31.

Zolnay, A. (2006). *Acoustic feature combination for speech recognition*, Universitätsbibliothek.