Investigation of Robust Speech Feature Extraction Techniques for Accents Classification of Malaysian English Speakers

by

Yusnita Mohd Ali
(1040610467)

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

School of Mechatronic Engineering
UNIVERSITI MALAYSIA PERLIS

2014

ACKNOWLEDGMENT

First and foremost, I thank God the Almighty, Allah Subhana Wata'ala, for giving His blessing to me to complete this fruitful journey of doctorate study. This research is never solitary and thus, I have many people to thank for making the writing of this thesis bearable and to reach the end of my battle successfully.

I should thank the Ministry of Higher Education Malaysia and my employer, Universiti Teknologi MARA (UiTM) for giving this golden opportunity and financial support, without which I may have not pursued this study diligently. Also, I would like to acknowledge the Vice Chancellor of University Malaysia Perlis (UniMAP), Brig. Jeneral Dato' Prof. Dr. Kamaruddin Hussin, the Dean and staffs of Mechatronic Engineering of UniMAP and the Dean and staffs of Centre for Graduate Studies of UniMAP for their support and encouragement throughout my candidature.

Especially, I would like to thank my main supervisor, Prof. Dr. Paulraj Murugesa Pandiyan for the three years and nine months of supervising my doctorate study. He has been very patient and dedicated with the writing and eventual completion of this thesis. His inputs and comments were tremendous and constructive and his insight proved to be very valuable. Towards the end, I voiced out to him my tiredness of the thesis writing, and it was him who reminded me of the effort and time I had invested in this research and the meaning of holding the doctorate title is just a license to start doing any research afterwards. And for that, I shall always be indebted to him.

My deepest gratitude also goes out to Prof. Dr. Sazali Yaacob, my co-supervisor especially during the first year of my candidature, who, despite his hectic schedule as prominent lecturer and deputy vice chancellor, graciously set aside time to entertain my fearful emotion of dark ages and guide me through the literature review of this study. I deeply appreciate his genuine concern for my well-being, his truly belief in me, his steady support and his kind words of encouragement.

I wish to also thank my second co-supervisor, Dr. Shahriman Abu Bakar for his advice and support especially in my data collection process and made the Malaysian English (MalE) accent database initiation most successful. Also, many thanks go to fellow members at the Intelligent Signal Processing Cluster of UniMAP, especially to my kind friends, Sathees Kumar Nataraj and Lim Sin Chee, who, closely share their knowledge and guide me very well through real examples.

It is no denying that I went through a very stressful time during my study. During these times of enormous difficulty, my family was the most that I sacrificed. My sincerest and utmost gratitude goes to my beloved husband, Syahreen bin Alim for his kindness, allegiance and unconditional love and support by taking finest care of our children, Harith Dzakwan and Ameera Mardhiyyah and for helping me with the household chores. I am also grateful to my beloved parents, Mohd Ali bin Yusof and Nik Jah binti Abdullah, for their patience, love, understanding and endless prayers towards my success.

TABLE OF CONTENTS

LIST OF TABLES

NO. **PAGE**

LIST OF FIGURES

NO. **PAGE**

LIST OF ABBREVIATIONS

AAC	Automatic Accent Classification
AE	American English
ANN	Artificial Neural Network
ANOVA	Analysis of Variance
ASR	Automatic Speech Recognition
BrE	British English
BruE	Brunei English
BS	Baseline
CALL	Computer-assisted Language Learning
CRD	Completely Randomized Design
CRs	Classification rates
dB	Decibels
DOE	Design of Experiments
DWT	Discrete Wavelet Transform
FF-MLP	Feed forward Multilayer Perceptron
FIR	Finite Impulse Response
FIS	Fuzzy Inference System
Formants	Formant frequencies
GMM	Gaussian Mixture Model
GSTs	Global Statistical Thresholds
HMM	Hidden Markov Model
Hz	Hertz
ICA	Independent Component Analysis
IWs	Isolated Words
KNN	K-nearest Neighbors
LDA	Linear Discriminant Analysis
LPC	Linear Prediction Coefficients
MalE	Malaysian English
MAP	Maximum-a-Posteori
MBSE	Mel-band Spectral Energy
MFCC	Mel-frequency Cepstral Coefficients
MLLR	Maximum Likelihood Linear Regression

MMS	Maximum-mean subtraction normalization
mse	Mean-squared errors
msec	milliseconds
MVN	Mean and variance normalization
PCA	Principal Component Analysis
PD	Pronunciation Dictionary
PPRLM	Parallel Phone Recognition Language Modeling
PRLM	Phone Recognition Language Modeling
QMF	Quadrature Mirror Filters
RFE	Recursive Feature Elimination
RP	Received Pronunciation
SBS	Statistical Band Selection
SFFs	Spectral Feature Fusions
SgE	Singapore English
SNR	Signal-to-Noise Ratio
STE	Short-time Energy
STFT	Short-time Fourier Transform
STs	Sentences
SVM	Support Vector Machine
SVM-RFE	Support Vector Machine-Recursive Feature Elimination
TTS	Text-to-Speech
V-UV	Voiced-Unvoiced
ZCR	Zero-crossing Rate

LIST OF SYMBOLS

α	Learning rate of ANN
β	Momentum rate of ANN

Kajian terhadap Teknik Penyarian Sifat Pertuturan Lasak untuk Pengelasan Loghat Bagi Penutur Bahasa Inggeris Berbangsa Malaysia

ABSTRAK

Sistem pengecaman pertuturan automatik (ASR) bukanlah suatu topik baru dalam pemprosesan pertuturan dan interaksi manusia-mesin. Ianya telah dikaji lebih daripada lima dekad lepas. Walau bagaimanapun, loghat kekal sebagai cabaran besar berkait rapat dengan kepelbagaian bahasa dalam isu-isu ASR masakini yang menggambarkan perbezaan pertuturan dalam sebutan dan intonasi seseorang yang mempunyai pelbagai perbezaan latar-belakang dari segi sosiolinguistik. Terdapat banyak dan pelbagai literatur yang telah mendedahkan kesan negatif daripada pelbagai loghat sebagai penyebab kemerosotan prestasi ASR. Walaupun loghat Bahasa Inggeris telah menjadi jenis loghat paling banyak dikaji kerana diangkat sebagai bahasa yang paling penting dan berprestij, MalE yang merupakan versi baru didalam 'New Englishes' dikalangan penutur bukan ibunda masih belum diterokai. Dalam produk pasaran ASR pada masa kini, MalE dianggap sebagai versi yang seragam secara konvensional walaupun tanggapan ini dipertikaikan oleh ramai sarjana dan penyelidik yang menganggap MalE sebagai penuturan yang terhasil daripada implikasi setempat kepelbagaian etnik. Kajian persepsi yang lepas telah melaporkan kemungkinan tinggi mengesan identiti etnik daripada penuturan Singapore English (SgE) dan Brunei English (BruE) yang boleh dijadikan perbandingan yang sesuai dengan MalE melalui ujian pendengaran. Pada masa ini, tiada kajian yang telah dilakukan untuk mengenal pasti asal usul etnik dari sampel penuturan MalE menggunakan pelbagai teknik analisis pertuturan dan algoritma pembelajaran mesin untuk pengelasan automatik yang lebih dapat dipercayai, standard dan tepat melalui kaedah eksperimen. Kajian ini merupakan satu cubaan untuk mengisi jurang tersebut dan untuk tujuan ini, satu pangkalan data baru loghat MalE telah dibina. Kajian ini merangsang sebutan jenis IWs dan STs daripada para pelajar universiti (lelaki dan perempuan) yang terdiri daripada tiga etnik utama di negara ini iaitu Melayu, Cina dan India yang mewakili para penutur berpendidikan tinggi menggunakan perkataan yang sensitif kepada loghat, dipilih daripada kajian yang lepas. Reka bentuk sistem yang dicadangkan terdiri daripada pra- pemprosesan, penyarian sifat dan pengelasan. Selain daripada pra-pemprosesan asas, kajian ini mencadangkan integrasi dengan sistem inferens kabur untuk segmentasi asas frem suara kepada bergetar-tidak bergetar (FIS V-UV) turut menyumbang kepada pelaksanaan sistem keseluruhan yang lebih baik berbanding sistem pengelasan loghat (AAC) konvensional. Satu kaedah baru yang dicadangkan yang dinamakan sebagai ambang statistik global (GSTs) untuk membina fungsi keahlian masukan-masukan tenaga pendek masa (STE) dan kadar lintasan sifar (ZCR) dalam segmentasi FIS V-UV telah mengurangkan jumlah frem yang perlu diproses di peringkat penyarian ciri. Keputusan eksperimen menunjukkan keberkesanan AAC-terbantu FIS V-UV yang dicadangkan menggunakan GSTs dengan peningkatan tertinggi dalam kadar ketepatan sebanyak 7.70% dan pengurangan frem sebanyak 24.26% berbanding AAC konvensional. Pada peringkat kedua, ciri-ciri akustik berkait rapat dengan loghat daripada tiga etnik dibangunkan melalui beberapa kaedah analisis bank-penuras, model saluran vokal, analisis hibrid dan analisis paduan. Daripada lapan vektor sifat yang diuji ke atas pangkalan data MalE, perihalan statistik tenaga spektrum jalur-Mel (MBSE), analisis komponen utama-berubah MBSE (disingkatkan sebagai PCA-MBSE), dua teknik hibrid ombak kecil diskret berubah diperolehi pekali ramalan

linear (disingkatkan sebagai DWT- LPC) dan dua paduan ciri spektrum (SFFs) diantara pekali cepstral frekuensi-Mel dan pekali ramalan linear dengan lima formants (disingkatkan sebagai MFCC-formants dan LPC-formants) adalah pendekatan baru dalam bidang ini. Keputusan eksperimen dari peringkat akhir mencadangkan bahawa teknik SFFs adalah pendekatan yang terbaik untuk pangkalan data ini untuk menghuraikan tiga loghat MalE dengan kadar ketepatan yang terbaik sebanyak 97.4%. Teknik ini telah mengatasi ciri-ciri MFCC standard sebanyak 7.8%. Di bawah analisis kekukuhan, teknik SFFs diikuti oleh PCA-MBSE telah menunjukkan kerintangan bunyi yang lebih baik daripada teknik penyarian sifat yang lain. Tesis ini juga menyumbang satu teknik baru pemilihan sifat yang dipanggil sebagai algoritma pemilihan jalur statistik (SBS) menggunakan kaedah yang mudah untuk memilih jalur, berdasarkan varians paling minimum terhadap skor di dalam kelas. Keputusan eksperimen menunjukkan bahawa SBS telah dapat meningkatkan prestasi AAC dengan mencapai kadar ketepatan yang lebih baik di antara 3.9% hingga 5.6%, keperluan memori yang kurang antara 22% kepada 55% dan kelajuan yang lebih cepat sebanyak 70% secara purata didalam masalah loghat tiga-kelas. Membandingkan tahap loghat antara jantina, kajian ini mencadangkan bahawa penutur lelaki mempunyai kadar loghat yabg lebih tinggi berikutan hasil yang konsisten kadar pengelasan yang lebih baik tidak kira apa-apa kaedah yang digunakan. Selain itu, kesimpulan dapat dibuat bahawa penuturan berterusan (STs) mempunyai kesan yang lebih baik daripada mod penuturan terpencil–perkataan (IWs).

Investigation of Robust Speech Features Extraction Techniques for Accents Classification of Malaysian English Speakers

ABSTRACT

Automatic speech recognition (ASR) system is not a new topic in speech processing and human-machine interaction. It has been established for more than five decades. However, accent remains a great challenge closely related to multilingualism in today's ASR issues which manifests speech differences in pronunciation and intonation of people from different sociolinguistics background. A large and growing body of literature has revealed the negative effects of various accents as impairment to the ASR performance. Although English accents have been the most studied accent varieties insofar as it is regarded the most important and prestigious international language, Malaysian English (MalE) which signifies a new variety within New Englishes of non-native speakers is still unexplored. In the ASR market product nowadays, conventional way is to treat MalE as a uniform variety despite this notion is disputed by many scholars and researchers who regard MalE as implication of localized ethnic speech diversity. Past perceptual studies have reported high possibility of detecting ethnic identities from Singapore English (SgE) and Brunei English (BruE) speech as appropriate comparator varieties to MalE accents using listening test setup. At present, no research has been done to identify ethnic origin from speech samples of MalE accented speech using multiple speech analysis techniques and machine learning algorithms for automatic classification for more reliable, standard and accurate experimental methods. This study is an attempt to fill that gap and for this purpose, a new database of MalE accents has been developed. The study elicits speech in isolated-words and continuous speech from university students of both genders of three main ethnics to represent educated speakers of Malay, Chinese and Indian groups using selected accent-sensitive words from previous studies. The design of the proposed system consists of pre-processing, feature extraction and classification stages. Apart from basic pre-processing, this study proposes integrating fuzzy inference system for voiced-unvoiced (FIS V-UV) frame basis segmentation by itself has contributed an improved overall implementation over conventional automatic accent classification (AAC) system. A new method is proposed, named as global statistical thresholds (GSTs) for establishing membership functions of short-time energy and zero crossing rate inputs in the FIS V-UV segmentation. This proposed segmentation has resulted in a reduced portion of speech activity to be taken further for feature extraction stage. The experimental results demonstrate the efficacy of the proposed FIS V-UV-assisted AAC using GSTs with the highest increase in accuracy rate of 7.70% and frame reduction rate of 24.26% over the conventional AAC. In the second stage, acoustic features correlated to accents of these three ethnics are developed through several techniques of filter bank analysis, vocal tract model, hybrid analysis and fusion analysis. Out of eight formulated feature vectors tested on the MalE database, statistical descriptors of Mel-band spectral energy (MBSE), principal component analysis-transformed MBSE (PCA-MBSE), two hybrid techniques of discrete wavelet transform-derived linear prediction coefficients (DWT-LPC) and two spectral feature fusions (SFFs) of popular Mel-frequency cepstral coefficients and linear prediction coefficients with five formants (MFCC-formants and LPC-formants) are new approaches in this field. The experimental results from the final stage suggest that SFFs techniques are the best approach for this database to classify the