



**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 Murugesu 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. Shahrizan 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

	PAGE
THESIS DECLARATION	i
ACKNOWLEDGMENT	ii
TABLE OF CONTENTS	iii
LIST OF TABLES	viii
LIST OF FIGURES	xii
LIST OF ABBREVIATIONS	xix
LIST OF SYMBOLS	xxi
ABSTRAK	xxii
ABSTRACT	xxiv
CHAPTER 1	1
INTRODUCTION	1
1.1 Chapter Overview	1
1.2 Research Background	1
1.3 Problem Statement	4
1.3.1 Perceptions of Accent in Attitudinal Study	4
1.3.2 Accent Challenges in Spoken Language Technology	5
1.3.3 Non-uniformity of Malaysian English Accents as a New Challenge in Automatic Speech Recognition	6
1.3.4 Lack of Robust and Salient Features and Integration with Efficient Pre-processing Techniques in Previous Studies	8
1.4 Research Objectives	9
1.5 Thesis Scope	10
1.6 Thesis Outline	11
CHAPTER 2	14
LITERATURE REVIEW	14
2.1 Introduction	14
2.2 The importance of Studying Malaysian English Accents	15
2.3 Accent Challenges to Automatic Speech Recognition Systems	21
2.4 Review on Automatic Accent Classification	21

2.4.1	Filter bank Analysis	22
2.4.2	Mel-frequency Cepstral Coefficients	23
2.4.3	Linear Prediction Coding	28
2.4.4	Formant Frequencies	31
2.4.5	Discrete Wavelet Transform	34
2.4.6	Previous Accent Classification Performance and Findings	38
2.5	Review on Speech Material	44
2.6	Review on Voiced-Unvoiced Segmentation of Speech	49
2.7	Significance of the Present Study	51
2.8	Summary	52
CHAPTER 3		54
RESEARCH METHODOLOGY		54
3.1	Introduction	54
3.2	Framework of the Study	54
3.3	Data Collection Procedure	57
3.3.1	Subjects/Participants	58
3.3.2	Types of Speech Material	60
3.3.3	Relevant Speech Material for Accent Elicitation	61
3.4	Malaysian English Accent Database	64
3.5	Summary	66
CHAPTER 4		68
PRE-PROCESSING AND FUZZY INFERENCE SYSTEM FOR VOICED-UNVOICED SEGMENTATION		68
4.1	Introduction	68
4.2	Basic Pre-processing Methods	68
4.2.1	Signal Normalization	69
4.2.2	Pre-emphasis Filtering	74
4.2.3	Frame-blocking, Overlapping Frames and Windowing	80
4.3	Fuzzy Inference System for Voiced-Unvoiced Frames Segmentation	85
4.3.1	Speech Model	86
4.3.2	Short-time Energy	88

4.3.3	Zero-crossing Rate	88
4.3.4	Design of Fuzzy Inference System	89
4.3.5	Algorithm and Implementation	100
4.4	Summary	105
CHAPTER 5		107
FORMULATION OF FEATURE VECTOR AND CLASSIFICATION OF ACCENT		107
5.1	Introduction	107
5.2	General Algorithm for Automatic Accent Classification	109
5.3	Design of Experiment for Statistical Analysis	111
5.4	Feature Extraction of Speech Signals	115
5.4.1	Mel-frequency Cepstral Coefficient	116
5.4.2	Statistical Descriptors of Mel-bands Spectral Energy	118
5.4.3	Algorithms for Linear Prediction Coefficients and Formants	122
5.4.4	Discrete Wavelet Transform-derived Linear Prediction Coefficients	125
5.5	Feature Reduction of High-dimensional Feature Vector	131
5.5.1	Principal Component Analysis	131
5.5.2	PCA-transformed Mel-bands Spectral Energy	133
5.6	Feature Selection of Important Features	136
5.7	Accent Classification	138
5.7.1	K-nearest Neighbors	140
5.7.2	Artificial Neural Network	143
5.8	Summary	146
CHAPTER 6		148
RESULTS AND DISCUSSION		148
6.1	Introduction	148
6.2	Analysis of Accent-sensitive Words in IWs Speech and Interaction Effect between Gender and Accent	149
6.2.1	Factorial Design on Accent-sensitive Words	149
6.2.2	Performance of Text-independent AAR	156
6.2.3	Performance of Text-dependent AAR	158

6.3	Analysis of Formants in STs Speech and Interaction Effect between Gender and Accent	159
6.4	Voiced-Unvoiced Segmentation using Fuzzy Inference System	163
6.5	Model and Evaluation Method for Automatic Accent Classification	168
6.5.1	K-nearest Neighbors	170
6.5.2	Artificial Neural Network	170
6.6	Mel-frequency Cepstral Coefficient-based Accent Classification	171
6.6.1	Varying K-parameter of KNN	172
6.6.2	Varying KNN Distance Metric	174
6.6.3	Varying MFCC Order	176
6.6.4	MFCC-based ANN Classifier	178
6.7	Statistical Descriptors of Mel-bands Spectral Energy-based Accent Classification and Feature Reduction using Principal Component Analysis	183
6.7.1	Varying K-parameter of KNN	184
6.7.2	Varying KNN Distance Metric	186
6.7.3	Varying Principal Components	187
6.7.4	MBSE-based ANN Classifier	188
6.8	Linear Prediction Coding-based Accent Classification	192
6.8.1	Varying K-parameter	192
6.8.2	Varying KNN Distance Metric	194
6.8.3	Varying LPC Order	195
6.8.4	LPC-based ANN Classifier	197
6.9	Discrete Wavelet Transform-derived Linear Prediction Coefficients-based Accent Classification	200
6.9.1	KNN Performance of Varying LPC Order in Hybrid DWT-LPC	201
6.9.2	ANN Performance of Fixed LPC Order	205
6.10	Spectral Feature Fusion	208
6.10.1	Spectral Feature Fusion-based KNN Classifier	209
6.10.2	Spectral Feature Fusion-based ANN Classifier	212
6.11	Feature Selection using Statistical Band Selection	215
6.11.1	Band Ranking using SBS	216
6.11.2	MBSE-based ANN Classifier with Applied SBS	218

6.11.3 MFCC-based ANN Classifier with Applied SBS	223
6.11.4 MBSE- and MFCC-based KNN Classifier with Applied SBS	227
6.12 Performance Comparison of Different Feature Vectors under Clean and Noisy Conditions	228
6.13 Summary	238
CHAPTER 7	242
CONCLUSION AND FUTURE RECOMMENDATION	242
7.1 Conclusion and Contribution	242
7.2 Future Recommendation and Limitation	246
REFERENCES	248
APPENDIX 3A	258
APPENDIX 3B	260
APPENDIX 3C	261
APPENDIX 6A	263
APPENDIX 6B	268
APPENDIX 6C	277
APPENDIX 6D	282
APPENDIX 6E	285
LIST OF PUBLICATIONS	287

LIST OF TABLES

NO.	PAGE
Table 2.1: Kandiah's (1998) framework of World Englishes. Adopted from Sharbawi (2009).	17
Table 2.2: Summary of the most significant studies of accent classification in chronological order spotting on problem, methodology and performance. .	38
Table 2.3: Summary of past studies regarding subjects/speech details.....	45
Table 3.1: Distribution of speakers in Malaysian English accent database in terms of ethnic group, gender and types of utterance.	65
Table 4.1: Mean and variance of the amplitude of several speech utterances using different methods of signal normalization.....	74
Table 4.2: Fuzzy rules to build voiced / unvoiced fuzzy inference system.....	97
Table 4.3: Resulted global statistical thresholds (GSTs) for STE and ZCR for building FIS input membership.	105
Table 5.1: Components of specification of the factorial design.	113
Table 5.2: Design of different resolution of Mel-filter bank.	116
Table 5.3: Mean and standard deviation of formant scores of the male speakers.	124
Table 5.4: Mean and standard deviation of formant scores of the female speakers.	124
Table 5.5: Proposed level dyadic division scale (LDDS) for dyadic DWT-LPC features.	127
Table 5.6: Contribution of LPC coefficients in multi-resolution sub-bands according to the rule of dyadic extraction of DWT-LPC features.....	130
Table 6.1: Results of ANOVA on different orders of MFCC scores (a) the 1 st and 2 nd order and (b) the 3 rd and 4 th order.	151
Table 6.2: Summary of significant results of individual isolated word across different MFCC-order for accent factor.	153
Table 6.3: Accuracy rates for text-independent AAC across different filter numbers..	156
Table 6.4: One-way ANOVA results summarizing the significant value and effect size of formants.....	161
Table 6.5: Significant results of post-hoc comparisons between three accent groups on formant scores using ANOVA.....	162
Table 6.6: Formant quartiles of the female speakers uttering the STs speech.....	163
Table 6.7: Formant quartiles of the male speakers uttering the STs speech.....	163

Table 6.8: Frame reduction rate for IWs speech and STs speech by employing FIS V-UV segmentation.	168
Table 6.9: Performance using 70-30 percentage of independent test samples on 12-MFCC-based KNN model across different K-parameter.	173
Table 6.10: Performance using 60-40 percentage of independent test samples on 12-MFCC-based KNN model across different K-parameter.	173
Table 6.11: Performance using 70-30 percentage of independent test samples on 12-MFCC-based KNN model using different distance metric.	176
Table 6.12: Performance using 60-40 percentage of independent test samples on 12-MFCC-based KNN model using different distance metric.	176
Table 6.13: Performance using 70-30 percentage of independent test samples using KNN model across different MFCC-order.	178
Table 6.14: Performance using 60-40 percentage of independent test samples using KNN model across different MFCC-order.	178
Table 6.15: Statistical classification rates across different order of MFCC for four test speech scenarios using maximum criterion for the output neuron state. ...	182
Table 6.16: Statistical classification rates across different order of MFCC for four test speech scenarios using threshold and margin criterion for the output neuron state.	182
Table 6.17: Performance using 60-40 percentage of independent test samples on 72-MBSE-based KNN model across different K-parameter.	185
Table 6.18: Performance using 60-40 percentage of independent test samples on 72-MBSE-based KNN model using different distance metric.	187
Table 6.19: Performance using 60-40 percentage of independent test samples using KNN model across different PCs.	188
Table 6.20: Statistical classification rates of MBSE and PCA-MBSE for four test speech scenarios using maximum criterion for the output neuron state.	190
Table 6.21: Statistical classification rates of MBSE and PCA-MBSE for four test speech scenarios using threshold and margin criterion for the output neuron state.	191
Table 6.22: Performance using 60-40 percentage of independent test samples on 12-LPC-based KNN model across different K-parameter.	194
Table 6.23: Performance using 60-40 percentage of independent test samples on 12-LPC-based KNN model using different distance metric.	195

Table 6.24: Performance using 60-40 percentage of independent test samples using KNN model across different LPC-order.	197
Table 6.25: Statistical classification rates across different order of LPC for four test speech scenarios using maximum criterion for the output neuron state. ...	199
Table 6.26: Statistical classification rates across different order of LPC for four test speech scenarios using threshold and margin criterion for the output neuron state.....	200
Table 6.27: Performance using 60-40 percentage of independent test samples using KNN model across different LPC order of dyadic DWT-LPC.	202
Table 6.28: Performance using 60-40 percentage of independent test samples using KNN model across different LPC order of uniform DWT-LPC.	202
Table 6.29: Statistical classification rates of 16-LPC and DWT-derived LPC for four test speech scenarios using maximum criterion for the output neuron state.	206
Table 6.30: Statistical classification rates of 16-LPC and DWT-derived LPC for four test speech scenarios using threshold and margin criterion for the output neuron state.....	207
Table 6.31: Statistical performance of spectral feature fusion of MFCC and formants across different orders of four speech test scenarios using KNN model....	211
Table 6.32: Statistical performance of spectral feature fusion of LPC and formants across different orders of four speech test scenarios using KNN model....	211
Table 6.33: Statistical classification rates of baseline features and spectral feature fusions using maximum criterion for the output neuron state.	214
Table 6.34: Statistical classification rates of baseline features and spectral feature fusions using threshold and margin criterion for the output neuron state..	214
Table 6.35: Selected bands based on statistical band selection for female dataset.....	217
Table 6.36: Selected bands based on statistical band selection for male dataset.....	217
Table 6.37: Performance of ANN accent classifier using baseline Mel-band spectral energy (BS-MBSE-72).....	219
Table 6.38: Performance of ANN accent classifier using statistical band selection (SBS-MBSE) – First stage.....	220
Table 6.39: Performance of ANN accent classifier using statistical band selection (SBS-MBSE) – Second stage.....	221

Table 6.40: Performance of ANN accent classifier using MFCC of baseline full bands (BS-MFCC13-72)	224
Table 6.41: Performance of ANN accent classifier using MFCC of statistical band selection(SBS-MFCC13) – First stage.....	224
Table 6.42: Performance of ANN accent classifier using MFCC of statistical band selection (SBS-MFCC13) – Second stage.....	225
Table 6.43: Performance drop percentage in Mean CR for different SNL level.	232
Table 6.44: Individual class classification rate (max CR) for different features in 40% test dataset of the best KNN model of IWs speech.....	238
Table 6.45: Individual class classification rate (max CR) for different features in 40% test dataset of the best KNN model of STs speech.....	238

© This item is protected by original copyright

LIST OF FIGURES

NO.	PAGE
Figure 2.1: Range of accents in Malaysia. Adopted from Gill (1993).....	18
Figure 2.2: Stylistic variation by ethnic group among educated Singaporean English speakers. Adopted from Deterding and Poedjosoedarmo (2000).....	20
Figure 2.3: Block diagram of MFCC feature extraction.	24
Figure 2.4: Mapping relationship of linear and Mel-frequency scales.....	25
Figure 2.5: Mel filter banks basis functions using 20 Mel-filters in the filter bank.	27
Figure 2.6: Block diagram of LPC parameters and formant frequencies.....	34
Figure 2.7: LPC Filter in (a) time domain; (b) frequency domain.....	34
Figure 2.8: Three-level wavelet decomposition tree in dyadic fashion.....	37
Figure 3.1: Conventional automatic accent classification (AAC) system.....	55
Figure 3.2: Accent-dependent ASR system with accent analyzer in the preceding stage.	56
Figure 3.3: Overall process flow of methodology used for the study of Male accents classification.	57
Figure 3.4: Variation in speaking mode and speaking style in human speech elicitation.	62
Figure 3.5: Speech recording setup and activities.....	66
Figure 4.1: Basic pre-processing of speech signal.	69
Figure 4.2: Signal waveform of a female Chinese speaker uttering “It would be better if a boy and a girl have more time for communication” before and after normalization.	71
Figure 4.3: Signal waveform of a male Chinese speaker uttering “Hello there. Your destination is in the east. Fifty-eight km from here” before and after normalization.	72
Figure 4.4: Histogram of a female Chinese speaker uttering “It would be better if a boy and a girl have more time for communication” before and after normalization.	73
Figure 4.5: Histogram of a male Chinese speaker uttering “Hello there. Your destination is in the east. Fifty-eight km from here” before and after normalization.	73
Figure 4.6: Frequency response of a single-tap high-pass finite impulse response (FIR) filter with different effect of the emphasis coefficient α	77

Figure 4.7: Speech waveforms of the original signal after mean-maximum mean subtraction normalization (MMS) and the pre-emphasized signal.	78
Figure 4.8: Close snapshots of between 0.9 sec and 1.0 sec of the (a) original signal and (b) pre-emphasized signal.....	79
Figure 4.9: Amplitude spectral plots of the (a) original signal and (b) pre-emphasized signal.	79
Figure 4.10: Time-domain characteristics of several commonly used window functions in signal processing to taper a signal to zero at the starting and ending of a frame.....	81
Figure 4.11: Concept of basic pre-processing of a speech signal using frame-blocking, overlapping and windowing.	84
Figure 4.12: Speech waveforms of the (a) pre-emphasized short-time frame signal and (b) windowed frame signal.	85
Figure 4.13: Source-filter model of human speech production.	87
Figure 4.14: Digital model of speech production following source-filter model.	87
Figure 4.15: Histograms to show distributions of (a) log STE and (b) ZCR of IWs.	90
Figure 4.16: Histograms to show distributions of (a) log STE and (b) ZCR of STs.	90
Figure 4.17: Design of voiced-unvoiced segmentation system using FIS engine based on STE and ZCR features.....	92
Figure 4.18: Settings for the membership functions of the IWs type FIS V-UV system for (a) STE input and (b) ZCR input.....	93
Figure 4.19: Settings for the membership functions of the STs type FIS V-UV system for (a) STE input and (b) ZCR input.....	94
Figure 4.20: Settings for membership functions of the speech output of the FIS V-UV system.....	95
Figure 4.21: Rule viewer editor to show fuzzy rules evaluation on a given input combination which resulted unvoiced-type of speech.	96
Figure 4.22: Rule viewer editor to show fuzzy rules evaluation on a given input combination which resulted voiced-type of speech.	96
Figure 4.23: Segmentation of frames into V-UV type based on characteristics of STE and ZCR using FIS.....	99
Figure 4.24: Examples of (a) a voiced frame (the 39 th frame) and (b) an unvoiced frame (the 77 th frame) using FIS V-UV system to justify the formulated fuzzy rules.	100

Figure 4.25: Steps taken in speech pre-processing to label voiced and unvoiced frames of the speech signal using FIS V-UV system.	101
Figure 4.26: Flowchart of fuzzy inference global statistical thresholds.....	102
Figure 4.27: Illustration of method used to calculate global statistical thresholds (GSTs).	103
Figure 5.1: Flowchart for the implementation of automatic accent classification (AAC).	110
Figure 5.2: Completely randomized design for 2-factor factorial design for any k^{th} -order of MFCC.....	113
Figure 5.3: The sequence of original standard order of observations in completely randomized design treatment cells.	114
Figure 5.4: Different approaches to generate features from accented speech.	116
Figure 5.5: Block diagram of MBSE feature extraction.....	119
Figure 5.6: Plot of a windowed frame signal and its LP spectrum of b13 the 65 th frame, the first utterance from Spk085, a Chinese female.	123
Figure 5.7: Block diagram of the proposed feature extraction of hybrid DWT-LPC approach using uniform-numbered LPC and dyadic numbered LPC extraction.	126
Figure 5.8: Structure of the combination using dyadic of LPC order from DWT sub-bands.....	128
Figure 5.9: Displaying the three-level decomposition signals of approximation and detail sub-bands of the word <i>bottom</i> (the 21 st frame of the first sample) from a Malay female speaker labeled as Spk018.	129
Figure 5.10: Displaying the original and reconstructed signals for the word <i>bottom</i> , the 21 st frame, 1 st sample from Spk018.....	129
Figure 5.11: Block diagram of PCA-MBSE feature extraction.....	134
Figure 5.12: Scatter plot for the first three statistical descriptors of the 1 st Mel-band in the original MBSE dataset.....	135
Figure 5.13: Scatter plot for the first three principal components of the PCA-MBSE transformed dataset.	136
Figure 5.14: Flowchart of K-nearest neighbors algorithm.	141
Figure 5.15: Process flow of classification phase using K-nearest neighbors.....	142
Figure 5.16: Architecture of two-layer feed-forward multilayer perceptron (FF-MLP) neural network.	143

Figure 5.17: Process flow of classification phase using artificial neural network.....	145
Figure 6.1: Normal plots of residual and factorial plots of main effects and interaction effects for the word <i>Aluminum</i> on 12 th order and the word <i>Bottom</i> on 5 th order MFCC.....	155
Figure 6.2: Performance of text-independent AAC using different Mel-resolution in MFCC features for (a) male speakers and (b) female speakers across different K-parameter.	157
Figure 6.3: Vocabulary ranking in terms of accent accuracy rates for male and female speakers.	159
Figure 6.4: Impact of performance of V-UV FIS-assisted AAC using GSTs on IWs speech for male and female speakers on 12-MFCC feature.....	165
Figure 6.5: Impact of performance of V-UV FIS-assisted AAC using GSTs on STs speech for male and female speakers on 12-MFCC feature.....	165
Figure 6.6: Improvement in accent recognition resulted from using voiced frames only for IWs speech of male and female speakers across different MFCC order.	167
Figure 6.7: Improvement in accent recognition resulted from using voiced frames only for STs speech of male and female speakers across different MFCC order.	167
Figure 6.8: Performance of KNN for (a) 70-30 partitions (b) 60-40 partitions across different K-parameter of four speech test scenarios.	174
Figure 6.9: Performance of KNN for (a) 70-30 partitions (b) 60-40 partitions across different distance metric of four speech test scenarios.	175
Figure 6.10: Performance of KNN for (a) 70-30 partitions (b) 60-40 partitions across different MFCC-order of four speech test scenarios.....	177
Figure 6.11: (a) Performance plots of best validation performance and (b) training states of AAC-ANN based on MFCC features.	180
Figure 6.12: Performance of ANN using (a) Max criterion (b) Threshold & Margin (TM) across different MFCC-order of four speech test scenarios.	181
Figure 6.13: Training time performance (epoch) of ANN for (a) IWs and (b) STs speech scenarios across different MFCC-order and genders.	183
Figure 6.14: Performance of test dataset on 72-MBSE of four speech test scenarios across different K-parameter.	185

Figure 6.15: Performance of the 72-MBSE dataset across different distance metric of KNN under four speech test scenarios.	186
Figure 6.16: Performance of PCA-MBSE test dataset across different dimension (PCs) of four speech test scenarios using KNN model.	188
Figure 6.17: Performance of ANN using (a) Max criterion (b) Threshold & Margin (TM) using MBSE-based and PCA-MBSE-based of four speech test scenarios.	190
Figure 6.18: Training time performance (epoch) of ANN of (a) IWs and (b) STs speech scenarios for MBSE and PCA-MBSE-based features.....	192
Figure 6.19: Performance of test dataset on 12-LPC of four speech test scenarios across different K-parameter.	193
Figure 6.20: Performance of the 12-LPC dataset across different distance metric of KNN under four speech test scenarios.	195
Figure 6.21: Performance of test dataset across different LPC-order of four speech test scenarios using KNN model.	196
Figure 6.22: Performance of ANN using (a) Max criterion (b) Threshold & Margin (TM) across different LPC-order of four speech test scenarios.....	199
Figure 6.23: Training time performance (epoch) of ANN of (a) IWs and (b) STs speech scenarios for LPC-based features.....	200
Figure 6.24: KNN Performance of conventional LPC and two methods of DWT-derived LPC across different order of IWs and STs speech modes.	203
Figure 6.25: Performance of the 16-LPC, 16-dyadic-DWT-LPC and 32-uniform DWT-LPC features of four speech test scenarios using KNN.	204
Figure 6.26: Performance of the 16-LPC, 16-dyadic-DWT-LPC and 32-uniform DWT-LPC features of four speech test scenarios using ANN.	206
Figure 6.27: Training time performance (epoch) of ANN of (a) IWs and (b) STs speech scenarios for LPC-based and DWT-derived LPC-based features.....	208
Figure 6.28: Performance of spectral feature fusion of MFCC and formants across different MFCC-order of four speech test scenarios using KNN model....	210
Figure 6.29: Performance of spectral feature fusion of LPC and formants across different LPC-order of four speech test scenarios using KNN model.	210
Figure 6.30: Performance of the baseline MFCC and LPC features and their spectral feature fusions of four speech test scenarios using KNN model.	212

Figure 6.31: Performance of ANN using (a) Max criterion (b) Threshold & Margin (TM) for the baseline MFCC and LPC and SFFs of four speech test scenarios.	213
Figure 6.32: Training time performance (epoch) of ANN of the IWs and STs speech scenarios for baseline features and spectral feature fusions.	215
Figure 6.33: Performance of the baseline features (BS-MBSE-72) and the statistical band selection methods i.e. SBS-MBSE-44 and SBS-MBSE-56 using ANN.	222
Figure 6.34: Performance of the baseline features (BS-MBSE) and the statistical band selection methods i.e. SBS-MBSE-32 and SBS-MBSE-40 using ANN.	223
Figure 6.35: Performance of the male and female speakers using baseline features and statistical band selected features on descriptors of MBSE and MFCC using ANN.	226
Figure 6.36: Performance of the baseline features (BS-MBSE) and the statistical band selection (SBS-MBSE) methods using KNN.	227
Figure 6.37: Performance of the male speakers using baseline features and statistical band selected features on descriptors of MBSE and MFCC using ANN and KNN.	228
Figure 6.38: Performance of the female speakers using baseline features and statistical band selected features on descriptors of MBSE and MFCC using ANN and KNN.	228
Figure 6.39: Robustness performance of various feature vectors under different noisiness levels for IWs speech of male speakers.	230
Figure 6.40: Robustness performance of various feature vectors under different noisiness levels for IWs speech of female speakers.	230
Figure 6.41: Robustness performance of various feature vectors under different noisiness levels for STs speech of male speakers.	231
Figure 6.42: Robustness performance of various feature vectors under different noisiness levels for STs speech of female speakers.	231
Figure 6.43: Performance (max CR) of different feature vector under clean and different SNR for IWs speech of male speakers.	233
Figure 6.44: Performance (max CR) of different feature vector under clean and different SNR for IWs speech of female speakers.	234

Figure 6.45: Performance (max CR) of different feature vector under clean and different SNR for STs speech of male speakers.235

Figure 6.46: Performance (max CR) of different feature vector under clean and different SNR for STs speech of female speakers.236

© This item is protected by original copyright

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-Posteriori
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

© This item is protected by original copyright

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 yang 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 terpercil-perkataan (IWPs).

© This item is protected by original copyright

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