COMPUTER-BASED MALAY SPEECH
ARTICULATION-PHONOLOGICAL ASSESSMENT SYSTEM

TING HUA NONG

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Doctor of Philosophy

Faculty of Electrical Engineering
Universiti Teknologi Malaysia

DECEMBER 2006
Dedicated to Jesus Christ,
my mum, dad, sister, brother,
my beloved wife,
my cute daughter, Joanne,
speech-language pathologists, and
children with articulation/phonological speech impairments.
ACKNOWLEDGEMENTS

I would like to thank my supervisor, Associate Professor Dr. Jasmy Yunus on his valuable guidance, enlightenments and insights. To me, he is a wonderful supervisor. Thanks for giving me flexibility and much patience in doing the research.

I would like to thank Ministry of Science, Technology and Innovation (MOSTI), Malaysia for sponsoring my study through National Science Fellowship (NST) and IRPA grant. Thanks for financially supporting me to present technical paper at 2005 IEEE Engineering in Medicine and Biology Conference (EMBC 2005), Shanghai, China. The research could not be a success without the funds.

Special thanks to Sandra Vandort, the speech-language pathologist at Department of Audiology and Speech Sciences, Universiti Kebangsaan Malaysia for her kind assistance and suggestions. Many thanks to Radziah Fahmy, Ernie Heliza Yusof and Zanira Haslinda Zakaria, for their assistance throughout the clinical trial as the speech-language pathologists at Department of Ear, Nose and Throat, Hospital Sultanah Aminah, Johor Bahru. Thanks to Sekolah Agama UTM Skudai for the collection of speech database as well as the control-subject trial. Thanks to Sekolah Kebangsaan Pendidikan Khas Johor Bahru for providing clients for the clinical assessment.

Warmest thanks to my beloved wife, Wong Suing Ming for her ever lasting support, patience and sharing.

Most of all, may thanksgiving, praise and glory be all to Jesus Christ, who gives me abundant faith, hope and love through the entire Ph. D. study. Amen.
ABSTRACT

Articulation and phonological disorders are associated with the problems of speech sound production and the improper usage of phonological rule in a language respectively. In a conventional assessment, client with articulation and/or phonological disorder is diagnosed manually by speech-language pathologist. Some problems of conventional assessment are identified such as subjectivity in transcription of speech sounds, no integration between the assessment tests of articulation evaluation, phonological evaluation, stimulability test and auditory discrimination test, expensive and non user-friendly instrumental devices and lack of computer-based Malay assessment. Thus, the study proposes an effective methodology for computer-based Malay articulation-phonological assessment in order to overcome the problems. The study also proposes three modular neural networks to recognize the Malay speech sounds in an objective way based on speech characteristics: place of articulation, manner of articulation and voicing. The training and testing of the neural networks involve 60 and 32 Malay children with a database size of 15,668 speech samples. The computer-based assessment requires a client to speak to a microphone, without any attached physical devices. All the results of speech sound transcription and analysis are summarized by the computer-based system. The proposed computer-based assessment system examines 22 Malay consonantal phonemes in isolated words, at three word positions: initial, middle and final. Clinical trial involving 14 Malay primary school children and three children with hearing and speech impairment reveals that the proposed computer-based assessment is valid and reliable. The computer-based system is able to achieve an accuracy of 65% in transcribing the speech sounds in a speaker-independent manner. Though the recognition rate is not very promising to make the system a total computer-based assessment, the proposed system is a helpful analysis tool for clinician, which is able to provide the results in a click of time.
ABSTRAK

# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>TITLE PAGE</td>
<td></td>
<td>i</td>
</tr>
<tr>
<td>DECLARATION</td>
<td></td>
<td>ii</td>
</tr>
<tr>
<td>DEDICATION</td>
<td></td>
<td>iii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td></td>
<td>iv</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td></td>
<td>v</td>
</tr>
<tr>
<td>ABSTRAK</td>
<td></td>
<td>vi</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td></td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td></td>
<td>xiv</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td></td>
<td>xviii</td>
</tr>
<tr>
<td>LIST OF SYMBOLS</td>
<td></td>
<td>xxi</td>
</tr>
<tr>
<td>LIST OF APPENDICES</td>
<td></td>
<td>xxv</td>
</tr>
</tbody>
</table>

1 INTRODUCTION

1.1 Introduction  
1.2 Overview of Articulation and Phonological Assessment  
1.3 Problem Statements  
1.4 Objectives  
1.5 Scope of the Work  
1.6 Outline of the Thesis  
1.7 Contribution of the Study  

1 1
2 ARTICULATION AND PHONOLOGICAL ASSESSMENT

2.1 Introduction 13
2.2 Malay Phonetics 13
2.3 Malay Phonology 19
  2.3.1 Distribution of Malay Vowels 19
  2.3.2 Distribution of Malay Dipthongs 20
  2.3.3 Distribution of Malay Consonants 21
2.4 Speech Disorders 22
  2.4.1 Voice Disorder 22
  2.4.2 Fluency Disorder 22
  2.4.3 Articulation and Phonological Disorder 23
2.5 Causes of Articulation/Phonological Disorder 25
2.6 Articulation and Phonological Assessment 26
  2.6.1 Articulation Assessment 27
    2.6.1.1 Templin-Darley Test of Articulation 29
    2.6.1.2 Photo Articulation Test 29
    2.6.1.3 Goldman-Fristoe Test of Articulation 30
    2.6.1.4 Arizona Articulation Proficiency Scale 30
    2.6.1.5 Fisher and Logemann Articulation Competence 31
  2.6.2 Phonological Assessment 31
    2.6.2.1 Compton-Hutton Phonological Assessment 36
    2.6.2.2 Assessment of Phonological Processes 37
    2.6.2.3 Phonological Process Analysis 37
    2.6.2.4 Natural Process Analysis 38
    2.6.2.5 Khan-Lewis Phonological Analysis 38
  2.6.3 Instrumental Assessment 39
    2.6.3.1 Computer-assisted Phonological Analysis 41
    2.6.3.2 Electromyograph 41
    2.6.3.3 Electropalatograph 42
2.6.3.4 Endoscopy
2.6.3.5 Spectrum Analyzer
2.6.3.6 Spectrograph
2.6.3.7 IBM SpeechViewer

2.7 Speech Recognition Technology
2.7.1 Speech Feature Extraction
   2.7.1.1 Formants
   2.7.1.2 Filter Bank Analysis
   2.7.1.3 Linear Predictive Coding
2.7.2 Speech Recognition Techniques
   2.7.2.1 Dynamic Time Warping
   2.7.2.2 Hidden Markov Model
   2.7.2.3 Artificial Neural Networks

2.8 Summary

3 MALAY ARTICULATION-PHONOLOGICAL ASSESSMENT SYSTEM
3.1 Introduction
3.2 Design Philosophy
3.3 Proposed Computer-based Assessment System
   3.3.1 Components of the Assessment
   3.3.2 Determination of Etiologies of Articulation/Phonological Disorder
3.3.3 Procedure of Computer-based Articulation-Phonological Assessment
   3.3.3.1 Speech Sample Collection
   3.3.3.2 Recording and Phonetic Transcription
   3.3.3.3 Articulation and Phonological Analysis
   3.3.3.4 Stimulability-Contextual Test
   3.3.3.5 Stimulability-Contextual Analysis
   3.3.3.6 Auditory Discrimination Test
   3.3.3.7 Auditory Discrimination Analysis
   3.3.3.8 Integrated Analysis
3.3.4 Formulation of Target Words for the Computer-based Assessment

3.3.4.1 Target Words for Articulation-Phonological Evaluation

3.3.4.2 Target Words for Stimulability-Contextual Test

3.3.4.3 Target Words for Auditory Discrimination Test

3.4 System Design

3.4.1 Speech Recognition System

3.4.1.1 Speech Database Collection

3.4.1.2 Phoneme Recognition

3.4.1.3 Off-line Experiments

3.4.1.4 On-line Experiments

3.4.2 Computer-based Malay Articulation-Phonological Assessment System

3.4.2.1 Articulation-Phonological Evaluation

3.4.2.2 Stimulability-Contextual Test

3.4.2.3 Auditory Discrimination Test

3.4.2.4 Analysis Results

3.5 Summary

4 PERFORMANCE OF PHONEME RECOGNIZER

4.1 Introduction

4.2 Singular Neural Networks

4.2.1 Performance of Initial SNN

4.2.2 Performance of Final SNN

4.2.3 Overall Performance of SNN

4.3 Modular Manner Neural Networks

4.3.1 Initial Modular Manner Neural Networks

4.3.1.1 Performance of Initial Affricates

4.3.1.2 Performance of Initial Fricatives

4.3.1.3 Performance of Initial Lateral-Trill
5.3.1 Validity Test 176
  5.3.1.1 Target Phonemes 176
  5.3.1.2 Conventional Malay Articulation/Phonological Test 178
  5.3.1.3 Computer-based Articulation-Phonological Assessment 181
5.3.2 Reliability Test 185
5.3.3 Computer Performance 187
5.4 Summary 188

6 CONCLUSIONS AND SUGGESTIONS 189
6.1 Conclusions 189
6.2 Suggestions 192
  6.2.1 Computer-based Malay Articulation-Phonological Assessment System 192
  6.2.2 Phoneme Recognition 193

REFERENCES 195

APPENDICES
APPENDIX A Result of Survey on Children’s Familiarity with Malay Words 208
APPENDIX B Articulation Error of Malay Children Between Seven and Ten Years Old 210
APPENDIX C Confusion Matrices of Phoneme Recognizers 211
APPENDIX D Publications 230
# LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE NO.</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>List of Malay vowels</td>
<td>14</td>
</tr>
<tr>
<td>2.2</td>
<td>List of pure Malay consonants</td>
<td>15</td>
</tr>
<tr>
<td>2.3</td>
<td>List of borrowed Malay consonants</td>
<td>15</td>
</tr>
<tr>
<td>2.4</td>
<td>Complete set of Malay consonants</td>
<td>16</td>
</tr>
<tr>
<td>2.5</td>
<td>Description of place of articulation</td>
<td>17</td>
</tr>
<tr>
<td>2.6</td>
<td>Description of manner of articulation</td>
<td>18</td>
</tr>
<tr>
<td>2.7</td>
<td>Description of voicing articulation</td>
<td>18</td>
</tr>
<tr>
<td>2.8</td>
<td>Distribution of Malay vowels</td>
<td>20</td>
</tr>
<tr>
<td>2.9</td>
<td>Distribution of Malay diphthongs</td>
<td>20</td>
</tr>
<tr>
<td>2.10</td>
<td>Distribution of Malay consonants</td>
<td>21</td>
</tr>
<tr>
<td>2.11</td>
<td>Types of Disfluency</td>
<td>23</td>
</tr>
<tr>
<td>2.12</td>
<td>Phonological processes of syllable structure changes</td>
<td>33</td>
</tr>
<tr>
<td>2.13</td>
<td>Phonological processes of substitution processes</td>
<td>34</td>
</tr>
<tr>
<td>2.14</td>
<td>Phonological processes of assimilation</td>
<td>35</td>
</tr>
<tr>
<td>3.1</td>
<td>List of target words</td>
<td>69</td>
</tr>
<tr>
<td>3.2</td>
<td>Number of phonemes at different word positions</td>
<td>71</td>
</tr>
<tr>
<td>3.3</td>
<td>Availability of phonemes at various word positions</td>
<td>75</td>
</tr>
<tr>
<td>3.4</td>
<td>Target words of Stimulability-Contextual Test</td>
<td>76</td>
</tr>
<tr>
<td>3.5</td>
<td>Target words and possible answers for the Auditory Discrimination Test</td>
<td>77</td>
</tr>
<tr>
<td>3.6</td>
<td>Summary of speech database collection</td>
<td>79</td>
</tr>
<tr>
<td>3.7</td>
<td>Number of tokens of training and test set</td>
<td>80</td>
</tr>
<tr>
<td>3.8</td>
<td>Number of tokens of training and test set for abnormal speech</td>
<td>81</td>
</tr>
</tbody>
</table>
3.9 Replacement of SFWW speech sounds with SIWW speech sounds 81
4.1 Performance of Initial SNN with different signal lengths 110
4.2 Common Errors of Initial SNN 112
4.3 Performance of Final SNN with different signal lengths 116
4.4 Common errors of the Final SNN 117
4.5 Performance of Initial Affricates with different signal lengths 120
4.6 Performance of Initial Fricatives with different signal lengths 121
4.7 Performance of Initial Lateral-Trill with different signal lengths 122
4.8 Performance of Initial Nasals with different signal lengths 123
4.9 Performance of Initial Plosives with different signal lengths 124
4.10 Performance of Initial Semivowels with different signal lengths 125
4.11 Performance of Final Fricatives with different signal lengths 126
4.12 Performance of Final Lateral-Trill with different signal lengths 127
4.13 Performance of Initial Nasals with different signal lengths 128
4.14 Performance of Initial Plosives with different signal lengths 129
4.15 Performances of Initial and Final MMNN 130
4.16 Performance of Integrated MMNN based on maximum recognition rate of individual classes 131
4.17 Performance of integrated MMNN based on maximum overall recognition rate 132
4.18 Selection of sub-networks of MMNN based on maximum recognition rate of individual class or overall recognition rate 133
4.19 Performance of hybrid integrated MMNN 134
4.20 Performance of Initial Alveolar with different signal lengths 135
4.21 Performance of Initial Bilabial with different signal lengths 136
4.22 Performance of Initial Glottal with different signal lengths 137
4.23 Performance of Initial Labio-Dental with different signal lengths 138
4.24 Performance of Initial Palatal with different signal lengths 139
4.25 Performance of Initial Velar with different signal lengths  140
4.26 Performance of Final Alveolar with different signal lengths  141
4.27 Performance of Final Bilabial with different signal lengths  142
4.28 Performance of Final Glottal with different signal lengths  143
4.29 Performance of LabioDental with different signal lengths  144
4.30 Performance of Final Palatal with different signal lengths  145
4.31 Performance of Final Velar with different signal lengths  146
4.32 Performances of Initial and Final MPNN  147
4.33 Performance of Integrated MPNN based on maximum recognition rate of individual classes  148
4.34 Performance of Integrated MPNN based on maximum overall recognition rate  149
4.35 Performance of Initial Voiced with different signal lengths  150
4.36 Performance of Initial Unvoiced with different signal lengths  151
4.37 Performance of Final Voiced with different signal lengths  152
4.38 Performance of Final Unvoiced with different signal lengths  153
4.39 Summary of Initial and Final MVNN  154
4.40 Performance of Integrated MVNN based on maximum recognition rate of individual classes  155
4.41 Performance of Integrated MVNN based on maximum overall recognition rate  156
4.42 Selection of MVNN based on maximum recognition rate of individual class or overall recognition rate  156
4.43 Performance of singular and modular neural networks  157
4.44 Performance of singular and modular neural networks without abnormal speech sounds  157
4.45 Performance of the singular and modular neural networks  158
4.46 Substitution errors of the singular and modular neural networks  160
4.47 Comparison between different phoneme recognizers  162
5.1 Overall performance of Malay children  166
5.2 Confusion Matrix of the articulation performance of Malay children  167
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.3</td>
<td>Summary of phonological processes</td>
<td>168</td>
</tr>
<tr>
<td>5.4</td>
<td>Confusion matrix of stimulability-contextual performance</td>
<td>169</td>
</tr>
<tr>
<td>5.5</td>
<td>Overall performance of Malay children</td>
<td>170</td>
</tr>
<tr>
<td>5.6</td>
<td>Confusion Matrix of the computer’s performance</td>
<td>172</td>
</tr>
<tr>
<td>5.7</td>
<td>Stimulability-Contextual performance of computer</td>
<td>174</td>
</tr>
<tr>
<td>5.8</td>
<td>Auditory discrimination performance of computer</td>
<td>175</td>
</tr>
<tr>
<td>5.9</td>
<td>Comparison between Ujian Fonologi Bahasa Melayu and MAPAS</td>
<td>177</td>
</tr>
<tr>
<td>5.10</td>
<td>List of target words used in the Ujian Fonologi Bahasa Melayu</td>
<td>179</td>
</tr>
<tr>
<td>5.11</td>
<td>Articulation scores obtained from Ujian Fonologi Bahasa Melayu</td>
<td>180</td>
</tr>
<tr>
<td>5.12</td>
<td>Articulation scores obtained from MAPAS</td>
<td>181</td>
</tr>
<tr>
<td>5.13</td>
<td>Consistency of speech production of Client 2</td>
<td>183</td>
</tr>
<tr>
<td>5.14</td>
<td>ADT scores and errors</td>
<td>184</td>
</tr>
<tr>
<td>5.15</td>
<td>Number of children involved in the inter-judge reliability test</td>
<td>185</td>
</tr>
<tr>
<td>5.16</td>
<td>Inter-judge reliability scores</td>
<td>186</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE NO.</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Mid-sagittal view of vocal tract, which shows anatomical place of articulation</td>
<td>17</td>
</tr>
<tr>
<td>2.2</td>
<td>Approaches of articulation-phonological assessment</td>
<td>45</td>
</tr>
<tr>
<td>2.3</td>
<td>Filter Bank Analysis</td>
<td>48</td>
</tr>
<tr>
<td>2.4</td>
<td>A MLP with one hidden layer</td>
<td>52</td>
</tr>
<tr>
<td>3.1</td>
<td>Proposed computer-based Malay articulation-phonological assessment</td>
<td>60</td>
</tr>
<tr>
<td>3.2</td>
<td>Classification of etiologies of articulation-phonological disorder</td>
<td>62</td>
</tr>
<tr>
<td>3.3</td>
<td>Procedure of computer-based articulation-phonological assessment</td>
<td>63</td>
</tr>
<tr>
<td>3.4</td>
<td>Description of word positions in a two-syllable word</td>
<td>70</td>
</tr>
<tr>
<td>3.5</td>
<td>Speech feature extraction</td>
<td>82</td>
</tr>
<tr>
<td>3.6</td>
<td>Architectures of SNN</td>
<td>83</td>
</tr>
<tr>
<td>3.7</td>
<td>Architectures of MMNN</td>
<td>84</td>
</tr>
<tr>
<td>3.8</td>
<td>Architectures of MPNN</td>
<td>85</td>
</tr>
<tr>
<td>3.9</td>
<td>Architectures of MVNN</td>
<td>86</td>
</tr>
<tr>
<td>3.10</td>
<td>GUI of computer-based Malay Articulation-Phonological Assessment</td>
<td>89</td>
</tr>
<tr>
<td>3.11</td>
<td>Sound recording during Articulation-Phonological Evaluation</td>
<td>91</td>
</tr>
<tr>
<td>3.12</td>
<td>Sound recording during Stimulability-Contextual Test</td>
<td>92</td>
</tr>
<tr>
<td>3.13</td>
<td>Auditory Discrimination Test</td>
<td>93</td>
</tr>
</tbody>
</table>
3.14 Words transcribed by SLP based on client’s pronunciation
3.15 Words transcribed by computer based on client’s pronunciation
3.16 Viewing analysis result of Performance using SLP mode (in percentage)
3.17 Viewing analysis result of Performance using SLP mode (in number of correct phoneme)
3.18 Viewing analysis result of Performance using Computer mode (in percentage)
3.19 Viewing analysis result of Performance using Computer mode (in number of correct phoneme)
3.20 Contrastic phone chart
3.21 Phonetic inventory of client based on SLP’s perception
3.22 Phonetic distribution
3.23 Articulation errors of client based on SLP’s perception
3.24 Phonological processes of client based on SLP’s transcription
3.25 Analysis result of Stimulability-Contextual Test based on SLP’s transcription
3.26 Analysis result of Stimulability-Contextual Test based on computer’s transcription
3.27 Summary of result of Auditory Discrimination Test
3.28 Summary of integrated analysis result
3.29 Classification of etiology of the articulation-phonology speech disorder
3.30 SLP’s comments
4.1 Training error vs epoch for Initial SNN with signal length of 130ms
4.2 Training error vs epoch for Initial SNN with signal length of 140ms
4.3 Recognition rate vs epoch for Initial SNN with signal length of 130ms
4.4 Recognition rate vs epoch for Initial SNN with signal length of 140ms

4.5 Training error vs epoch for Final SNN with signal length of 170ms

4.6 Recognition rate vs epoch for Final SNN with signal length of 170ms

5.1 Waveform of “cendawan”, which was detected as two-syllable sound

5.2 Waveform of “dewan”, which was detected as one-syllable sound

5.3 Waveform of “selipar”, which was detected as two-syllable sound
LIST OF SYMBOLS

ADT  Auditory Discrimination Test
ak   LPC coefficients
α    Momentum term
ANN  Artificial Neural Network
BP   Back-propagation
bhj  Biases at the hidden layer
byk  Biases at the output layer
C    Consonant
Č    Phoneme /c/
cm   Cepstral coefficients
CBST Computer-based Speech Training
CV   Consonant-Vowel
CVC  Consonant-Vowel-Consonant
δj   Error information term at hidden layer
δk   Error information term at output layer
Δbhj Bias correction term at hidden layer
Δbyk Bias correction term at output layer
Δwj  Weight correction term between input layer and hidden layer
Δwkj Weight correction term between hidden layer and output layer
dB   Decibel
DTW  Dynamic Time Warping
δ    Phoneme /dh/ or /dz/
Emin Minimum error or global minimum
Ep   Mean square error for a single input pattern
<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_{\text{rms}}$</td>
<td>Root mean square error</td>
</tr>
<tr>
<td>e(n)</td>
<td>Prediction error</td>
</tr>
<tr>
<td>E(n)</td>
<td>Mean squared error used in autocorrelation method</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>Mid-high central vowel e</td>
</tr>
<tr>
<td>EF</td>
<td>Error Function</td>
</tr>
<tr>
<td>EGG</td>
<td>Electroglottograph</td>
</tr>
<tr>
<td>$E_p$</td>
<td>Mean square error for single pattern</td>
</tr>
<tr>
<td>$E_{\text{rms}}$</td>
<td>Root mean square error</td>
</tr>
<tr>
<td>EMG</td>
<td>Electromyograph</td>
</tr>
<tr>
<td>EPG</td>
<td>Electropalatograph</td>
</tr>
<tr>
<td>F1</td>
<td>Formant 1</td>
</tr>
<tr>
<td>F2</td>
<td>Formant 2</td>
</tr>
<tr>
<td>F3</td>
<td>Formant 3</td>
</tr>
<tr>
<td>F4</td>
<td>Formant 4</td>
</tr>
<tr>
<td>F5</td>
<td>Formant 5</td>
</tr>
<tr>
<td>G</td>
<td>Gain of the excitation</td>
</tr>
<tr>
<td>GUI</td>
<td>Graphics User Interface</td>
</tr>
<tr>
<td>HMM</td>
<td>Hidden Markov Models</td>
</tr>
<tr>
<td>HNN</td>
<td>Hidden Neuron Number</td>
</tr>
<tr>
<td>IBM</td>
<td>International Business Machines</td>
</tr>
<tr>
<td>IPA</td>
<td>International Phonetic Alphabets</td>
</tr>
<tr>
<td>ISPA</td>
<td>Interactive System for Phonological Analysis</td>
</tr>
<tr>
<td>$\ddot{\text{j}}$</td>
<td>Phoneme /j/</td>
</tr>
<tr>
<td>$\ddot{\text{q}}$</td>
<td>Plosive-Glottal sound</td>
</tr>
<tr>
<td>$\dddot{\text{x}}$</td>
<td>Phoneme /kh/</td>
</tr>
<tr>
<td>LIPP</td>
<td>Logical International Phonetic Programs</td>
</tr>
<tr>
<td>LPC</td>
<td>Linear Predictive Coding</td>
</tr>
<tr>
<td>LR</td>
<td>Learning Rate</td>
</tr>
<tr>
<td>MAPAS</td>
<td>Malay Articulation-Phonological Assessment System</td>
</tr>
<tr>
<td>MLP</td>
<td>Multi-layer Perceptron</td>
</tr>
<tr>
<td>MMNN</td>
<td>Modular Manner Neural Networks</td>
</tr>
<tr>
<td>MPNN</td>
<td>Modular Place Neural Networks</td>
</tr>
<tr>
<td>Symbol</td>
<td>Term</td>
</tr>
<tr>
<td>--------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>ms</td>
<td>Milliseconds</td>
</tr>
<tr>
<td>MSE</td>
<td>Mean square error</td>
</tr>
<tr>
<td>MVNN</td>
<td>Modular Voicing Neural Networks</td>
</tr>
<tr>
<td>( \eta )</td>
<td>Learning rate</td>
</tr>
<tr>
<td>( \eta )</td>
<td>Phoneme /ng/</td>
</tr>
<tr>
<td>( \eta )</td>
<td>Phoneme /ny/</td>
</tr>
<tr>
<td>PCC</td>
<td>Percentage of Consonants Correct</td>
</tr>
<tr>
<td>R(n)</td>
<td>Autocorrelation Function</td>
</tr>
<tr>
<td>RR</td>
<td>Recognition Rate</td>
</tr>
<tr>
<td>SCT</td>
<td>Stimulability-Contextual Test</td>
</tr>
<tr>
<td>SFWW</td>
<td>Syllable-Final-Word-Within</td>
</tr>
<tr>
<td>SFWF</td>
<td>Syllable-Final-Word-Final</td>
</tr>
<tr>
<td>SIWI</td>
<td>Syllable-Initial-Word-Initial</td>
</tr>
<tr>
<td>SIWW</td>
<td>Syllable-Initial-Word-Within</td>
</tr>
<tr>
<td>SLP</td>
<td>Speech-language pathologist</td>
</tr>
<tr>
<td>SLTRU</td>
<td>Speech and Language Therapy Research Unit</td>
</tr>
<tr>
<td>SNN</td>
<td>Singular Neural Network</td>
</tr>
<tr>
<td>s(n)</td>
<td>Speech samples</td>
</tr>
<tr>
<td>( \hat{s}(n) )</td>
<td>Predicted speech samples</td>
</tr>
<tr>
<td>( $ )</td>
<td>Phoneme /sy/</td>
</tr>
<tr>
<td>f</td>
<td>Phoneme /sy/</td>
</tr>
<tr>
<td>TDNN</td>
<td>Time Delay Neural Network</td>
</tr>
<tr>
<td>TFL</td>
<td>Total Frame Length</td>
</tr>
<tr>
<td>TFN</td>
<td>Total Frame Number</td>
</tr>
<tr>
<td>( \theta )</td>
<td>Phoneme /th/</td>
</tr>
<tr>
<td>UFBM</td>
<td>Ujian Fonologi Bahasa Melayu</td>
</tr>
<tr>
<td>UKM</td>
<td>Universiti Kebangsaan Malaysia</td>
</tr>
<tr>
<td>UTM</td>
<td>Universiti Teknologi Malaysia</td>
</tr>
<tr>
<td>u(n)</td>
<td>Normalized excitation</td>
</tr>
<tr>
<td>V</td>
<td>Vowel</td>
</tr>
<tr>
<td>VOT</td>
<td>Voice Onset Time</td>
</tr>
<tr>
<td>( W_{ij} )</td>
<td>Weights between input layer and hidden layer</td>
</tr>
<tr>
<td>( W_{\text{initial}} )</td>
<td>Initial weights</td>
</tr>
<tr>
<td>Symbol</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
</tr>
<tr>
<td>( w_{jk} )</td>
<td>Weights between hidden layer and output layer</td>
</tr>
<tr>
<td>( W_{\text{local}} )</td>
<td>Weights during local minimum</td>
</tr>
<tr>
<td>( W_{\text{min}} )</td>
<td>Weights during global minimum</td>
</tr>
<tr>
<td>WI</td>
<td>Weight Initialization</td>
</tr>
<tr>
<td>( w(n) )</td>
<td>Window function</td>
</tr>
<tr>
<td>( w_m )</td>
<td>Weighting window for cepstral coefficients</td>
</tr>
<tr>
<td>( x_i )</td>
<td>Input neuron of the Multi-layer Perceptron</td>
</tr>
<tr>
<td>x</td>
<td>Phoneme /gh/ or [gh] sound</td>
</tr>
<tr>
<td>( X_n )</td>
<td>Bandpass filter</td>
</tr>
<tr>
<td>( \ddot{z} )</td>
<td>Phoneme /zh/ or [zh] sound</td>
</tr>
</tbody>
</table>

- Presence

- Absence
# LIST OF APPENDICES

<table>
<thead>
<tr>
<th>APPENDIX</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Result of survey on children’s familiarity with Malay words</td>
<td>208</td>
</tr>
<tr>
<td>B</td>
<td>Articulation errors of Malay children between seven and ten years old (transcription by SLP)</td>
<td>210</td>
</tr>
<tr>
<td>C</td>
<td>Confusion matrices of phoneme recognizers</td>
<td>211</td>
</tr>
<tr>
<td>D</td>
<td>Publications</td>
<td>230</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

1.1 Introduction

In 1970s, while speech-language pathologist (SLP) was still investigating conventional method in assessing articulation and phonological disorder manually, it was really hard to think of a computer that could recognize and understand speech of human beings. It was a dream of SLP that computer could transcribe the speech sounds and analyze the result for them.

It was an impossible mission in the past. However, with the advancement of computer technology in its computing speed and memory, nowadays, the fiction-like mission could be a reality! How could this impossible task become possible in the assessment of articulation and/or phonological disorder? Is computer really able to understand the human speech sounds? How good can the computer transcribe the human speech sounds? Is the computer’s performance comparable to human perception? Lots of questions could be asked before a computer-based system is to be developed. We may doubt the ability and capability of the computer in transcribing and understanding the human speech sounds. Thus, it is hoped that the study could provide some insightful answers to the above questions.
1.2 Overview of Articulation and Phonological Assessment

Early traditional assessment focuses primarily on the articulation disorders. This is because the disorder is regarded exclusively as the result of motor act. The focus of the traditional assessment is on the individual speech sounds. Examples of traditional assessment of articulation are Templin-Darley Test of Articulation (Templin and Darley, 1969), Fisher and Logemann Test of Articulation (Fisher and Logemann, 1971), Photo Articulation Test (Pendergest, Dickey, Selmar and Sudar, 1984), Arizona Articulation Proficiency Scale and Goldman-Fristoe Test of Articulation (Goldman and Fristoe, 1986). The tests can be administered in 15 minutes or up to 45 minutes, depending on the experience and skill of the SLP as well as the severity of the articulation disorders. These articulation assessments are easy and fast to administer. Besides that, they provide quantifiable list of misarticulated sounds in different word positions and standardized scores. However, most of these tests test one target phoneme in single words, ignoring the rests of the phonemes including consonants and vowels in the phonetic context. In addition, most of these tests examine the speech sounds in a limited number of phonetic contexts. More importantly, most of these tests do not detect the patterns of error in client’s speech, thus provides insufficient information about the client’s phonological system.

Due to the failure of the articulation assessment to provide the information regarding the linguistic usage of the speech sounds, thus SLP began to examine the speech production problems from the perspectives of phonological theory in 1970s. Two approaches of phonological assessment were introduced: distinctive feature analysis and phonological process analysis. Distinctive feature analysis attempts to specify the characteristics of phonemes – according to the presence (+) and absence (–) of each feature that distinguishes or contrast one speech sound from another. Distinctive feature analysis is limited of its application due to its time-consuming administering time, complexity and its binary functionality. On the other hand, phonological process analysis exits as the more common phonological assessment. Phonological process analysis describes the systematic sound change or simplification of adult speech patterns in terms of phonological processes. Examples of phonological process analysis are Assessment of Phonological Processes (Hodson, 1986), Assessment Link Between Phonology and Articulation (Lowe, 1986),
Phonological Process Analysis (Weiner, 1979) and Compton-Hutton Phonological Assessment (Compton & Hutton, 1978). The major drawback of these phonological assessments is that the time required to administer and analyze is much longer if compared to the articulation assessment.

Articulation assessment and phonological assessment are referred to as conventional assessment or perceptual assessment. The conventional assessment requires a diagnostician or SLP to manually conduct the assessment including picture-naming, tape-recording, transcribing speech sounds, form-recording, scoring, analyzing, interpreting and summarizing. Single words, phrases, sentences or conversation can be used to elicit the speech samples from the client. These perceptual assessments are subjective in nature. The subjectivity is much dependent on the skill, experience, knowledge and expertise of the SLP and it differs from one diagnostician to another.

McGuire (1995) points out that the advent of computers and the overall improvement in instrumentation have allowed the practitioner to quantify various speech dimensions and study the results in relation to perceptual observations. The instrumental assessment can be described in three ways: computer-assisted phonological analysis, acoustical approach and physiological approach.

Haynes and Pindzola (1997) describe the beauty of the computer-assisted phonological analysis is that the SLP does not have to spend several more hours of organizing data, scanning the transcript over and over again, and performing mathematical operations. In addition, the computer also provides elegant summaries of the data such as phonetic inventories, measures of severity and even suggested treatment targets on some programs. In other words, the computer-assisted analysis saves lots of time for SLP and provides greater efficiency. However, Haynes and Pinzola (1997) point out that these computer analysis do take away much of tedious work but not all. The SLP still have to obtain speech sample manually, transcribe the sample, input the sample into the computer through keyboard, and in many cases do some other work, responding to menus and prompts produced on the screen. Examples of the computerized or computer-assisted phonological analysis are Computer Analysis of Phonological Processes (Hodson, 1985), Computerized
Physiological assessment is carried out as aerodynamic measurement and movement measurement of vocal tract during speech or/and non-speech tasks such as pneumatachograph, spirometer, electromyograph (EMG), electropalatograph (EPG) and endoscopy. According to Buder et al. (1996), the physiological methods can provide fairly direct indications of speech and voice processes, but they tend to encumber the speaker with sensing devices, often give only a restricted view of the processes of interest and sometimes are difficult to correlate with perceptual or acoustic evaluations. In addition, the device needed for some physiological studies can be expensive and not easily accommodated to subjects of different ages and sizes.

Acoustical approach investigates common acoustic features such as frequency, amplitude and duration of speech sounds. Common and widely used acoustic measurement devices include spectrum analyzer and spectrograph. The advantage of acoustic measures over the perceptual measure is that it can provide quantified and objective evaluation. The major drawback of these acoustic analyses is that the displayed information is difficult to be understood by the client. Only those SLP with special training can analyse and interpret the results. These acoustic devices are very useful in analysing the speech signals but may not be efficient in terms of time if used as the assessment tool. Many user-friendly speech software systems are developed based on acoustical approach such as IBM SpeechViewer and Kay’s Elecmetrics programmes. However, these systems are used as training tools rather than as diagnostic tools and they are mostly available in foreign language such as English.

Physiological and acoustical approaches do provide means of quantified and objective evaluations in some ways. However, most of the physiological and acoustical devices are analysis devices, which can be used in the session of
assessment and therapy. These devices are still dependent on the SLP to interpret the results, thus introducing the issue of subjectivity.

Haynes and Pindzola (1997) suggest an ideal computer system, where the client simply talks into a microphone that is plugged into a computer and in a few seconds a printout appears that reveal the secrets of the phonological system. Unfortunately, the current instrumental assessments are still unable to achieve that and most of jobs in the assessment need to be done manually.

1.3 Problem Statements

The problems of traditional articulation and phonological assessment as well as the instrumental assessment are identified and summarized into five problem statements as below:

1) Conventional articulation and phonological assessment and other tests such as stimulability test and auditory discrimination test work independently towards describing the speech disorders. Each of the tests addresses different diagnosis of the speech problem. Articulation assessment examines client from the aspect of motor-based speech production while phonological assessment assesses client from the perspective of speech error patterns. Stimulability test evaluates the stimulability of the error sounds and auditory discrimination test examines the phonological processing or speech perceptual skills of the client. Each of the assessments has its own limitations. More importantly, there is no integration between these conventional assessments. Thus, there is a need to develop a computer-based assessment system to integrate these assessments and inter-relate the analysis of the different assessments into a more comprehensive analysis result. The integration of these results makes it possible to determine the etiologies of the speech disorder, which the individual tests can not solve with. According to Crystal (1981), children with different etiologies can be classified into three groups based on the phonetics and phonology: (1) Normal phonology, but abnormal phonetics (speech production), (2) Normal phonetics, but abnormal phonology, and (3) Abnormal phonology and phonetics. The integration could make a clear-cut to determine the actual and specific
client’s problems either articulatory or phonological, which result in appropriate treatment strategies to remediate the client’s problem.

2) Conventional articulation and phonological assessment are subjective in nature especially in the tasks of obtaining speech samples, transcribing speech samples, analyzing and interpreting the results, defining levels of severity and planning for the therapy. The subjectivity is much dependent on the skill, experience, knowledge and expertise of the SLP. This issue of subjectivity differs from one diagnostician to another. Thus, there is a need to develop a computer-based system to assess the articulation-phonological speech disorder in an objective way by utilizing computer phoneme recognition technology to transcribe the speech samples and define the levels of severity.

3) Despite the instrumental assessment providing many advantages, they do have problems and limitations. Firstly, computerized phonological analysis systems do help SLP save a lot of time doing analysis, but SLP still have to obtain speech samples manually, transcribe the speech sample and input the data into computer for analysis. Secondly, the physiological instrument such as EPG, pneumatachograph and spirometer are expensive and these devices are to be attached to the client. Furthermore, these devices only examine certain aspect of interest per device. The acoustic approach such as spectrum analyzer and spectrograph are reliable but the diagrams are difficult to be understood by the client. Besides that these devices may not appropriate to be used as assessment tools because they require much more time to administer. On the other hand, most of the instrumental devices are used as speech training tool rather than assessment tool. Thus, a computer-based methodology needs to be proposed to overcome the limitation and problems of current instrumental assessment devices. The proposed methodology needs to be able to i) to take away most of manual assessment tasks and further save much more time for SLP, ii) examine the client in a more natural way without any external circuit to be connected to the client, iii) to develop a computer-based system that is user-friendly to the SLP and client, where the client can understand the diagrams and analysis result more easily.
4) So far there is no standard Malay articulation-phonological assessment in Malaysian clinics and hospitals. There are some clinician derived tests available in clinics and hospital such as Azizah’s Phonological Test (Nordin, 1999). Furthermore there is no computer-based articulation-phonological assessment in Malaysia. Currently, most of the computer-based system are available in foreign languages especially English. These foreign-language-based computer-based systems are basically speech training system rather than assessment system. Thus, there is a need to develop a Malay-language computer-based system to be used in Malaysian clinics and hospitals and hopefully a standard computer-based Malay articulation-phonological assessment.

5) Most of the current research of Malay speech recognition primarily focuses on the adult speech recognition at isolated word level and it is very limited at phoneme level. These isolated word speech recognition systems, however are not feasible to be implemented in the computer-based speech articulation-phonological assessment system to recognize Malay consonantal phonemes of Malay children. Thus, there is a need to develop a phoneme recognition system that is able to recognize Malay consonantal phonemes of Malay children in a speaker-independent manner.

1.4 Objectives

The objectives of the study are

(1) To develop an effective methodology for a Malay articulation-phonological assessment by integrating different assessment tests into a comprehensive computer-based assessment system.

(2) To develop a prototype computer-based Malay articulation-phonological assessment system to assess speech problem among Malay children in an objective and user-friendly way.

(3) To propose new neural network architectures in recognizing Malay speech sounds in a speaker-independent manner.
1.5 Scope of the Work

The objectives of the study can be achieved through these scopes of study. The scopes clearly define the specific area of interest of the study.

Speech disorders are categorized into three major groups: voice disorder, fluency disorder and articulation/phonological disorder. Each of the disorders concerns different aspects of the speech problem, indicated by their names. The study focuses only on the aspects of articulation/phonological disorder, excluding the rest of speech disorders.

The articulation/phonological disorders are addressed in two major aspects: assessment and therapy. Assessment is used to diagnose the disorder whereas therapy is used to remediate the disorder based on the assessment result. The study investigates a computer-based methodology for the articulation/phonological assessment. The study does not look into the therapy aspect of the speech problem.

The articulation/phonological disorder can be a speech problem for children or adults, or both. The study investigates the speech problem of children rather than adult’s. Children are definitely in the process of learning adult’s speech, thus facing more problems than adults. Malay children between seven and ten years old are selected for the study.

The articulation/phonological disorder can exist in any language such as English, Chinese, Tamil and Malay. Each of the languages has its own unique phonetic and phonological system. As a result, the articulation/phonological disorder can be different from one language to another. The study only investigates the computer-based articulation/phonological assessment of Malay language among Malay children.

During the articulation/phonological assessment, the speech samples can be elicited from the children in terms of single words, phrases, sentences or conversation. The study investigates the use of single words to elicit the speech samples from the children, which is easier and faster. These single words can consist
of single or multiple syllables. Three types of syllable structures are available in the study: consonant-vowel (CV), vowel-consonant (VC) and consonant-vowel-consonant (CVC).

Malay speech sounds are basically comprised of phonemes of vowel, diphthongs and consonants. The study only investigates computer-based articulation-phonological assessment of Malay consonantal phonemes. Consonants are selected rather than vowels because children with articulation/phonological disorder have primarily problems with consonants. 22 Malay consonants are selected for the study. The study does not investigate the consonants that are low in frequency of occurrence such as /q/, /gh/, /kh/, /th/, /dh/ and /zh/. The study examines the assessment of consonants in all word position such as initial, middle and final positions. The study does not look into the consonant clusters or blends.

There are quite a number of speech feature extraction techniques available to extract the speech features such as formants, filter banks and Linear Predictive Coding (LPC). The study adopts LPC in extracting the speech feature because it is easy to implement and fast to compute.

The study uses Artificial Neural Networks in recognizing the Malay consonantal speech sounds. A phoneme recognizer is developed using Multi-layer Perceptron (MLP) with one hidden layer to recognize the Malay consonantal sounds in speaker-independent manner. The MLP is trained with stochastic Back-propagation training, where weights of the MLP are updated after presentation of each training pattern.

Lastly, the study uses software development tools of Microsoft Visual C++ and C++ programming languages to develop the computer-based Malay speech articulation-phonological assessment system as well as the phoneme recognition system.
1.6 Outline of the Thesis

Chapter 2 simply describes the literature review of the articulation and phonological assessments. First of all, Malay phonetics and phonology are introduced to provide a brief information on the Malay speech sounds such as consonants and vowels. Speech disorders and causes of the disorders are then discussed. Subsequently, the chapter discusses in length about the conventional articulation and phonological assessment as well as the instrumental approaches. Lastly, the speech recognition technology such as speech feature extraction and recognition techniques are reviewed. The speech recognition system is used as the front-end processor of the computer-based assessment system to transcribe the speech sounds.

Chapter 3 illustrates the system design of the proposed computer-based Malay Articulation-Phonological Assessment System (MAPAS). Procedure of the articulation-phonological assessment is described. The formulation of target words for Articulation and Phonological Evaluation, Stimulability-Contextual Test and Auditory Discrimination Test is discussed. Lastly, the chapter presents the system design of the speech recognition system including proposed modular neural networks as well as the MAPAS.

Chapter 4 presents the performance of the speech recognition system – phoneme recognizer in a speaker-independent manner through offline experiments. The phoneme recognizer is implemented using Linear Predictive Coding as the speech feature extractor and Multi-layer Perceptron (MLP) as the recognizer. The performances of different architectures of neural networks are evaluated including the singular neural networks and modular neural networks. Three modular neural networks are proposed based on the speech characteristics: Modular Manner of Articulation, Modular Place of Articulation and Modular Voicing. The effect of speech signal length on the performance of phoneme recognizer and the ability of these neural networks in recognizing abnormal speech sounds are investigated.

Chapter 5 deals with the online or real-time performance of the computer-based Malay Articulation-Phonological Assessment System (MAPAS). The clinical
trials are carried out in primary school and hospital involving normal and abnormal Malay children between seven and ten years old. The validity and reliability of MAPAS is evaluated as well. The comparison between conventional articulation assessment and computer-based system is discussed.

Chapter 6 simply concludes the study and provides suggestions for future development of the study. The conclusions of the study include the efficiency of MLP in recognizing the Malay consonantal phonemes and abnormal speech sounds in speaker-independent way, computer-based Malay articulation-phonological assessment system and its ability to recognize Malay consonantal phonemes and abnormal speech sounds in speaker-independent way. The suggestions are based on the limitations of the current computer-based assessment system and the phoneme recognizer.

1.7 Contribution of the Study

The study is solving the problems and limitations of conventional articulation and phonological assessments. The existing conventional articulation and phonological assessments, stimulability test and auditory discrimination test work independently and separately to describe the articulation/phonological disorders. Each of the tests looks at different perspective of the speech problem and thus has their own strengths and limitations. The study has proposed a new effective methodology by integrating all these assessments and tests into a computer-based assessment to diagnose the speech problem more efficiently from different perspectives. The integration of these analysis results makes it possible to determine the etiologies of the articulation/phonological disorder, which the individual tests of the conventional assessment can not solve. Through the proposed methodology, the etiologies of the articulation/phonological disorder can be categorized into four groups: normal articulation and phonology, normal articulation and abnormal phonology, abnormal articulation and normal phonology, abnormal articulation and phonology.
Articulation and phonological assessments are still conducted manually by SLP in most Malaysian clinics and hospitals. Currently, there is no standard Malay assessment test available and most of tests are informal clinician-derived tests. Moreover, there is no computer-based Malay articulation/phonological assessment system available in Malaysia. The study has proposed a new computer-based method to diagnose Malay articulation/phonological disorder among Malay children in an objective way and in a faster way. The proposed computer-based assessment system is called Malay Articulation-Phonological Assessment System (MAPAS), which is able to provide analysis of result in a click of time.

Current research of Malay speech recognition has been carried out comprehensively at isolated word level, but very limited at phoneme level. These studies focus on the recognition of adult speech sounds. The existing Malay speech recognition systems of adult speech sounds at isolated word level are not feasible to be implemented to recognize the Malay phonemes in computer-based speech articulation-phonological assessment. Thus, the study has proposed new architectures of neural networks in recognizing 22 Malay consonantal phonemes of Malay children in a speaker-independent manner. The architectures are proposed in modular neural networks based on speech characteristics: place of articulation, manner of articulation and voicing. The proposed modular architectures perform better than conventional singular neural networks. The performance of the proposed modular neural networks is comparable with other phoneme recognizers of foreign languages. Besides that, the study also investigates the effect of different speech signal lengths on the performance of the neural networks. Most of the phoneme recognitions, however examine the speech signal length in a fixed length. The study concludes that the performance of the neural networks is very dependent on the appropriate length of speech signal.
References


