AUTOMATIC SPEECH RECOGNITION MODEL FOR DYSLEXIC CHILDREN READING IN BAHASA MELAYU

HUSNIZA HUSNI

DOCTOR OF PHILOSOPHY (Ph.D.)
UNIVERSITI UTARA MALAYSIA
2010

Kolej Sastera dan Sains
(UUM College of Arts and Sciences)
Universiti Utara Malaysia

PERAKUAN KERJA TESIS / DISERTASI
(Certification of thesis / dissertation)

Kami, yang bertandatangan, memperakukan bahawa
(We, the undersigned, certify that)

HUSNIZA HUSNI

calon untuk Ijazah **PhD**
(candidate for the degree of)

telah mengemukakan tesis / disertasi yang bertajuk:
(has presented his/her thesis / dissertation of the following title):

"AUTOMATIC SPEECH RECOGNITION MODEL FOR DYSLEXIC CHILDREN READING IN BAHASA MELAYU"

seperti yang tercatat di muka surat tajuk dan kulit tesis / disertasi.
(as it appears on the title page and front cover of the thesis / dissertation).

Bahawa tesis/disertasi tersebut boleh diterima dari segi bentuk serta kandungan dan meliputi bidang ilmu dengan memuaskan, sebagaimana yang ditunjukkan oleh calon dalam ujian lisan yang diadakan pada : ***26 April 2010.***
That the said thesis/dissertation is acceptable in form and content and displays a satisfactory knowledge of the field of study as demonstrated by the candidate through an oral examination held on:
April 26, 2010.

Pengerusi Viva: *(Chairman for Viva)*	**Prof. Dr. Zulkhairi Md Dahalin**	Tandatangan *(Signature)*
Pemeriksa Luar: *(External Examiner)*	**Assoc. Prof. Dr. Sharifah Mastura Syed Mohamad**	Tandatangan *(Signature)*
Pemeriksa *Luar*: *(External Examiner)*	**Assoc. Prof. Dr. Manjit Singh Sidhu**	Tandatangan *(Signature)*
Nama Penyelia/Penyelia-penyelia: *(Name of Supervisor/Supervisors)*	**Assoc. Prof. Dr. Zulikha Jamaludin**	Tandatangan *(Signature)*

Tarikh:

(Date) ***April 26, 2010***

AUTOMATIC SPEECH RECOGNITION MODEL FOR DYSLEXIC CHILDREN READING IN BAHASA MELAYU

A Thesis submitted to the UUM College of Arts and Sciences in full fulfillment of the requirements for the degree of Doctor of Philosophy
Universiti Utara Malaysia

by
Husniza Husni

PERMISSION TO USE

IKHTISAR

Kanak-kanak disleksia mengalami masalah pembelajaran yang berkait rapat dengan sistem fonologi yang mengganggu perkembangan mereka dalam kemahiran membaca dan mengeja. Bagi mereka, membaca dan mengeja merupakan proses yang sukar, meletihkan dan kurang menarik perhatian. Konsentrasi yang lemah juga mengakibatkan mereka sebolehnya mengelak untuk belajar membaca dan mengeja. Masalah disleksia ini menyebabkan mereka melakukan kesalahan pembacaan walaupun ketika membaca dan mengeja perkataan-perkataan mudah. Walau bagaimanapun, keadaan ini tidak bermakna mereka memilik tahap IQ yg rendah berbanding kanak-kanak normal. Selalunya, kanak-kanak disleksia memiliki tahap IQ yang setara dengan rakan-rakan mereka yang lain dan mungkin juga lebih tinggi. Fakta ini merumuskan bahawa kanak-kanak disleksia mempunyai potensi jika diberikan bantuan dan sokongan padu seperti motivasi dan kaedah pengajaran yang sesuai. Dengan kemajuan teknologi dalam bidang pendidikan, aplikasi berasaskan komputer dilihat dapat membantu proses pengajaran dan pembelajaran khususnya bagi membaca dan mengeja. Justeru itu, kajian ini merupakan inisiatif yang mencadangkan sebuah model *automatic speech recognition* (ASR) yang berupaya 'mendengar' dan seterusnya mengenalpasti kesalahan bacaan bagi kanak-kanak ini. Skop kajian menghadkan kepada pemodelan dan pengecaman perkataan terpilih dalam Bahasa Melayu (BM) yang terkandung di dalam silibus pengajaran tahap satu di sekolah rendah. Untuk mengesyorkan pemodelan ASR dalam BM, sebuah model pembacaan khusus bagi kanak-kanak disleksia terlebih dahulu dicadangkan. Teknik kajian etnografi, iaitu pemerhatian dan temubual secara tidak formal, telah diguna untuk mendapatkan kesalahan-kesalahan bacaan yang juga melibatkan kesalahan ejaan. Sepuluh orang murid berusia 7 hingga 14 tahun yang mempunyai tahap kebolehan membaca yang hampir sama telah dipilih menyertai kajian. Mereka terdiri daripada murid di dua buah sekolah rendah yang menawarkan kelas khas disleksia. Sebanyak 6112 bacaan telah direkod dalam bentuk audio dan daripada itu, sejumlah 6051 kesalahan bacaan telah berjaya dikenal pasti. Antara kesalahan yang paling menonjol adalah kesalahan berkaitan penggantian huruf vokal, penyingkiran huruf konsonan, kesalahan berkaitan dengan huruf nasal dan penggantian konsonan. Justeru itu, model ASR tersebut telah mengambilkira aspek kesalahan bacaan yang paling kerap tersebut dan menjadikan ia sebagai elemen penting bagi tujuan pengecaman. Model ASR tersebut mengambilkira kesalahan bacaan bagi sesuatu perkataan sebagai alternatif kepada sebutan yang betul di mana kesalahan-kesalahan itu telah dimodelkan ke dalam model leksikalnya. Strategi penambahbaikan fonem juga digunakan bagi tujuan meningkatkan ketepatan pengecaman, iaitu menurunkan kadar kesalahan perkataan (*word error rate*, WER). Untuk itu, sebuah prototaip pengecam (*recognizer*) telah dihasilkan bagi membolehkan proses penilaian dilakukan terhadap ketepatan pengecam yang berasaskan model cadangan. Penilaian ketepatan ini diukur menggunakan WER dan kadar pengesanan kesalahan bacaan (*miscue detection rate*, MDR) yang berkait rapat dengan *false alarm rate* (FAR). Pengecam berasaskan model cadangan tersebut berjaya mencapai tahap kepuasan WER serendah 25% dan 80.77% untuk MDR dengan kadar FAR 16.67%.

Katakunci: *Pengecaman suara automatik, model leksikal, kanak-kanak disleksia, Bahasa Melayu.*

ABSTRACT

Dyslexic children suffer from dyslexia, a condition that profoundly impedes reading and spelling ability due to its phonological origin. Often, these children found reading and spelling difficult, exhaustive, and less interesting, and thus they are self-withdrawn from the learning process. When reading and spelling, they make many mistakes even for simple, common words that they themselves found embarrassing. However, this does not mean that they have lower IQ level than normal children. In fact, dyslexic children have average or high level of IQ and thus have a lot of potential when given the right help and support such as motivational support and suitable teaching techniques. With advancement in technology in education, computer-based applications are used to stimulate the learning process of reading and spelling. Hence, this study is an initiative towards proposing an automatic speech recognition (ASR) model to enable computer to 'listen' should incorrect reading occurs. The scope of this study focuses on modeling and recognizing single Bahasa Melayu (BM) words within the school syllabus for level one (*tahap satu*) dyslexic pupils of primary schools. To propose the ASR model, a reading and spelling model of dyslexic children reading in BM is first proposed, which models reading at word recognition level. To propose such model, ethnographic techniques are employed namely informal interviews and observation, in order to obtain the reading and spelling error patterns of dyslexic children. A number of ten dyslexic children, aged between 7 to 14 years old whose reading level is similar, participated in the study. These children are recruited from two public schools that offer special dyslexia classes for the children. A total of 6112 utterances are recorded in audio form resulting in a total of 6051 errors of various types. Among these, the patterns that are most frequently made by these children are of 'Substitutes vowel', 'omits consonant', 'nasals', and 'substitutes consonant'. The ASR model is proposed taking into consideration the error patterns that make lexical model a fundamental element for speech recognition. The lexical model is modeled to treat mispronunciations as alternative pronunciations or variants of target words. To that, a phoneme refinement strategy is applied aiming to increase recognition accuracy. A prototype recognizer is developed based on the proposed model for further evaluation. The evaluation is performed to evaluate the recognizer's performance in terms of accuracy, measured in word error rate (WER) and miscue detection rate (MDR) that is closely related to false

alarm rate (FAR). The recognizer scores a satisfying 25% of WER and a relatively high MDR of 80.77% with 16.67% FAR.

Keywords: *Automatic speech recognition, lexical model, dyslexic children, Bahasa Melayu.*

ACKNOWLEDGEMENTS

'Alhamdulillah', praise be to Allah, The Most Beneficent, The Most Merciful.

First and foremost, my deepest thank you goes to my supervisor, Assoc. Prof. Dr Zulikha Jamaludin for without her support, guidance, and help this thesis would not have been successfully materialized. I cannot fully express my gratitude to the exceptional advice every time I seek enlightenment, be it regarding PhD or about life itself. So, thank you.

My sincere appreciation to the important people who have contributed, in various ways, by providing the help and support needed during the research: Dr. Abd. Razif b. Abd. Razak, consultant paediatrician of Kedah Medical Center; Dr. Tan of Bureau of Learning Difficulties (BOLD), Penang; Pn. Sariah Amirin, presiden of Dyslexia Association Malaysia, Kuala Lumpur; Kim Wilson-Rymer of Davis Dyslexia Program, Ontario, Canada for continuous information on dyslexia; headmasters and dyslexia teachers of Sekolah Kebangsaan Taman Tun Dr. Ismail 2 (SKTTDI 2), Kuala Lumpur, especially Datin Zahrah Abdulllah, Dyslexia programme coordinator of SKTTDI 2 and Madam Ratana and Madam Bibi of Sekolah Kebangsaan Jalan Datuk Kumbar, Alor Setar.

A special thank you goes to Assistant Professor John-Paul Hosom of Center of Spoken Language Understanding (CSLU), Department of Science and Engineering, School of Medicine, Oregon Health and Science University, N.W. USA for his generous assistance especially in modelling the ASR lexical, developing, and testing the ASR recognizer.

Last but not least, a heartiest thank you to my husband, Suhairizam Omar, and my son, Humayl Suhair for everything. Thank you also to my parents, Husni and Norain, and to my friends for their love and continuous pray, support, and understanding. Without them, the PhD journey would not have been wonderful. Thank you.

TABLE OF CONTENTS

PERMISSION TO USE i

IKHTISAR ii

ABSTRACT iv

ACKNOWLEDGEMENTS vi

TABLE OF CONTENTS vii

LIST OF TABLES x

LIST OF FIGURES xii

LIST OF ACRONYMS xv

NAMING CONVENTION xiv

CHAPTER 1: INTRODUCTION

1.0 Introduction 1
1.1 Research Taxonomy 5
1.2 Motivation 6
1.3 Problem Statement – Challenges at Hand 7
1.4 Research Objective 10
1.5 The Scope 12
1.6 Contribution of Research 15
1.7 Overall Thesis Summary 17
1.7.1 Research Recapitulated 17
1.7.2 Thesis Organization 19

CHAPTER 2: LITERATURE REVIEW

2.0 Introduction 21
2.1 Using ASR Technology to Facilitate Reading: Theory-based Perspective 23
2.2 The Vocabulary and Age-Matched Dyslexic Children 26
2.3 Reading and Spelling Error Patterns 28
2.4 Dyslexic Children's Reading Model 31
2.5 ASR 1, 2, 3 37
2.5.1 ASR for BM 37
2.5.2 Discrete vs. Continuous ASR System for Dyslexic 38
2.5.3 ASR Methods 42
2.5.3.1 Template Matching 42
2.5.3.2 Acoustic-Phonetic Approach 43
2.5.3.3 Hidden Markov Model 44
2.5.3.4 Artificial Neural Network Approach 45
2.5.3.5 Hybrid Method of HMM/ANN 46
2.5.4 ASR Models 48
2.5.5 Lexical Model and Context-dependent Modeling 52

2.6 Evaluation: The Recognition Accuracy 54
2.7 Chapter Summary 58

CHAPTER 3: METHODOLOGY

3.0 Introduction 60
3.1 Task 1: Data Collection 62
3.1.1 The vocabulary 64
3.1.2 The participants – Dyslexic children 68
3.1.3 Primary data collection 71
3.2 Task 2: Model Creation 74
3.2.1 Data analysis 75
3.2.2 Pronunciation model 79
3.2.3 Improvement of the reading model 81
3.2.4 Improvement of the ASR model 84
3.3 Task 3: ASR Recognizer Development 89
3.4 Task 4: Evaluation 92
3.4.1 The WER evaluation 92
3.4.2 MDR and alignment technique 93
3.5 Summary 95

CHAPTER 4: THE DEVELOPMENT OF THE ASR RECOGNIZER

4.0 Introduction 97
4.1 Setting the 'Stage' 98
4.1.1 The speech files (.wav) 99
4.1.2 The transcription files (.txt) 100
4.1.3 The phoneme files (.phn) 101
4.2 Creating Descriptions Files 103
4.2.1 The lexical model as the key component in ASR 103
4.2.2 The corpora 109
4.2.3 The grammar and parts specifications 110
4.2.4 The `.info` files 112
4.3 The Training Process 114
4.4 Phoneme Refinement Strategy 125
4.5 Summary 128

CHAPTER 5: EVALUATION: WORD ERROR & MISCUE DETECTION

5.0 Introduction 129
5.1 Evaluation and Results: WER 130
5.2 Evaluation: MDR 135
5.3 Results: MDR 137
5.4 Analysis and Discussion 138
5.4.1 The Effects of Phoneme Refinement on WER, MDR, and FAR 139
5.4.2 The Overall Performance 142
5.4.3 The Shortcoming 144
5.5 Summary 145

CHAPTER 6: CONCLUSION & FUTURE DIRECTIONS

6.0 Introduction 147
6.1 Research Recapitulation 148
6.2 The Objectives Achieved 150
6.2.1 To collect vocabulary in BM 152
6.2.2 To recognize reading and spelling patterns 153
6.2.3 To model dyslexic children's reading in BM 154
6.2.4 To model an ASR for the reading model 155
6.2.5 To develop an ASR recognizer based on the ASR model 156
6.2.6 To evaluate in WER and MDR 157
6.3 Upshot – The reading model and the ASR model 157
6.4 Future works 160
6.5 Summary 162

REFERENCES 163

APPENDICES

Appendix A: Consonant–Vowel (CV) pairs in syllable patterns of BM words 178
Appendix B: The Readings of Dyslexic Children According to Word Syllable Patterns 180
Appendix C: Percentage od Occurences of Modeled Pronunciations 204
Appendix D: Algorithm for transcription and miscue detection in tcl 219
Appendix E: CSLU Toolkit's Training Instructions 223
Appendix F: Target words and hypotheses (incorrect recognition). 238
Appendix G: Place and Manner of Articulation for BM 240
Appendix H: The Alignment Technique (`wordsfa2net.28`) 241
Appendix I: The Alignment Technique (`worsfanet.10`) 246
Appendix J: The Alignment Technique 251

LIST OF RELATED PUBLICATION TO THESIS 261

LIST OF TABLES

Table 1.1.	Research problem and research questions addressed in this study together with summary of methods and expected deliverables.	18
Table 2.1	Phonological error types and their definitions and example words adapted from Sawyer et al. (1999).	30
Table 2.2.	The different effects of discrete ASR and continuous ASR adapted from Higgins and Raskind (2000).	40
Table 2.3.	Discrete vs. continuous ASR effects on various skills adapted from Higgins and Raskind (2000).	41
Table 2.4.	The WER percentage of recognition tasks using HMM and HMM/ANN performed in various domains.	48
Table 2.5.	MDR and FAR of various ASR.	58
Table 3.1.	Example of clusters of common words within level 1 syllabus, tabled accordingly in random cluster sampling technique.	66
Table 3.2.	Examples of words for each of the syllable patterns of consonant-vowel categories and used as stimuli for data collection.	67
Table 3.3.	The readings of *sunyi*, which have been transcribed into correspondence spellings according to how they were pronounced.	75
Table 3.4.	Classification of errors for read words, which have been transcribed into corresponding spellings.	78
Table 3.5.	Categories of error types for analysis of read words error classification and the corresponding definitions.	79
Table 3.6.	Reading and spelling errors, in descending percentile, of dyslexic children reading aloud controlled vocabulary in BM.	80
Table 3.7.	Pronunciation models for word *ibu*, *aku, dan, rumah,* and *kereta* in Worldbet.	82
Table 3.8.	Alignment of target, transcript, and hypothesis tokens for miscue detection.	96
Table 4.1.	The relationship of speech file, transcription file, and phoneme file.	101
Table 4.2.	The mispronunciations of *wad* and their frequencies that later	108

suggest future potential feedback.

Table 4.3. The commands used in training the recognizer and their functions. 123

Table 4.4. Target words and recognized words (hypotheses) comparison. 129

Table 4.5. The letters and their pronunciation variations allowed in the lexical model to enhance accuracy. 130

Table 5.1. The output of `select_best.tcl` on development dataset, extracted from the summary file showing word level accuracy. 135

Table 5.2. The output of `select_best.tcl` on development dataset after force-alignment, extracted from the summary file showing word level accuracy. 136

Table 5.3. The output of `select_best.tcl` on development dataset after second force-alignment, extracted from the summary file showing word level accuracy. 136

Table 5.4. The output of `select_best.tcl` on development dataset after third force-alignment, extracted from the summary file showing word level accuracy. 137

Table 5.5. The best network for the development dataset fed into cycles of force- alignment. 138

Table 5.6. The final output evaluated on test dataset. 138

Table 5.7. The alignment of transcripts, targets, and hypotheses. 141

Table 5.8. The output of `select_best.tcl` on development dataset. 143

Table 5.9. The output of `select_best.tcl` on test dataset. 144

Table 5.10. The performance of the proposed lexical model and its counter model measured in WER, MDR, and FAR. 145

Table 5.11. The performance of the proposed recognizer among the recognizers built for similar purposes. 147

Table 6.1. The results and deliverables and their contributions outlined for the specific objective achieved. 156

LIST OF FIGURES

Figure 2.1.	Basic architecture of the dual-route model adopted from Coltheart et al. (2001).	33
Figure 2.2.	The structure of the Seidenberg-McClelland's computer simulated model of reading, adopted from Seidenberg and McClelland (1989).	34
Figure 2.3.	A simple model of cognitive processes involved in word recognition of a single word, adopted from Ellis (1993).	36
Figure 2.4.	HMM structure consists of five nodes with their transitions and recursive transitions adopted from Rabiner and Juang, 1993.	45
Figure 2.5.	A multilayer neural network, illustrating an input layer, a hidden layer, and an output layer.	47
Figure 2.6.	The general components involved in speech recognition.	50
Figure 2.7.	HMM architecture for isolated word recognition (source: Rabiner and Huang (1993).	51
Figure 2.8.	The illustration of the hybrid HMM/ANN architecture adopted from Shobaki et al. (2000).	52
Figure 3.1.	The methods and deliverables encapsulated in the research framework.	63
Figure 3.2.	The process of selecting the words to serve as vocabulary and stimuli.	69
Figure 3.3.	DAS's causal model of dyslexia, illustrating the various environmental factors in a bilingual or multilingual community, adopted from DAS (2003).	71
Figure 3.4.	A reading model that models the most frequent reading and spelling errors as a result of dyslexic children reading aloud controlled vocabulary in BM.	84
Figure 3.5.	The reading model and how it supports the ASR model.	87
Figure 3.6.	The proposed ASR model with inclusion of reading errors as active lexicon and phoneme refinement.	89
Figure 3.7.	The process flow of ASR recognizer development.	93

Figure 3.8.	Running `select_best.tcl` to obtain WER on test dataset.	94
Figure 4.1.	Example of speech files recorded displayed in SpeechView.	102
Figure 4.2.	The phonetic labels (in Worldbet) of every phoneme of the word *kelapa* and their positions given in millisecond (ms).	104
Figure 4.3.	The view of speech window (top), 2-D spectrogram window (middle), and the label window (bottom).	105
Figure 4.4.	A snippet of the lexicon file containing target words' pronunciations as well as mispronunciations represented in Worldbet, separated by the OR (|) operator.	107
Figure 4.5.	Alternative pronunciations of target words modeled into the lexical model.	111
Figure 4.6.	The context-dependent phoneme model of the word *aku*.	114
Figure 4.7.	2-D Spectrogram of phoneme `A` in the word *bawang*.	115
Figure 4.8.	Speech files divided into three datasets according to the 3:1:1 ratio.	116
Figure 4.9.	The overview of the training process for developing a prototype speech recognizer using CSLU Toolkit.	118
Figure 4.10.	Automatic generations of training, development, and testing datasets using `find_files.tcl` command.	119
Figure 4.11.	An illustration of the 3 layer feed-forward network architecture.	120
Figure 4.12.	A snippet of the output of the vector file created showing the maximum number of categories on the left column, which gives 77.	120
Figure 4.13.	The results for learn rate and total errors while training the hybrid network using `nntrain.exe` of CSLU Toolkit.	122
Figure 4.14.	The flow of training process in detail.	124
Figure 4.14 (cont.).	The flow of training process in detail (continues).	125
Figure 4.15.	The force-alignment process.	127
Figure 4.15 (cont.).	The force-alignment process (continues).	128
Figure 4.16.	The illustration of phoneme refinement rule applied to modeling the lexicon.	130

Figure 5.1.	The final performance comparison between different lexical models of the same data, measured in WER and MDR-FAR.	145
Figure 6.1.	The theoretical foundation constitutes to dyslexia and phonetically similar vocabulary and how it benefits the ASR.	154

LIST OF ACRONYMS

3M	*Membaca, Menulis, Mengira* (Reading, Writing, and Calculating)
ADD	Attention Deficit Disorder
ADHD	Attention Deficit Hyperactivity Disorder
ANN	Artificial Neural Network
ASCII	American Standard Code for Information Interchange
ASR	Automatic Speech Recognition
BM	Bahasa Melayu
CoLiT	Colorado Literacy Tutor
CSLU	Center for Spoken Language Understanding
C-V	Consonant-Vowel
DAS	Dyslexia Association Singapore
DC	Dyslexic Children
FAR	False Alarm Rate
HMM	Hidden Markov Model
HMM/ANN	Hybrid of Hidden Markov Model and Artificial Neural Network
ICT	Information and Communication Technology
IPA	International Phonetic Alphabet
LD	Learning Disability
LISTEN	Literacy Innovation that Speech Technology ENables
LLP	Liberated Learning Project
MDR	Miscue Detection Rate
MoE	Ministry of Education
NICHCY	National Dissemination Center for Children with Disabilities
OGI KIDS	Oregon Health & Science Univeristy (OGI) Children's Corpus (called KIDS Corpus)
PDP	Parallel Distribute Processing
PIPP	*Pelan Induk Pembangunan Pendidikan*
R&D	Research and Development
RAD	Rapid Application Development
TTS	Text-to-Speech
UTMK	*Unit Terjemahan Melalui Komputer*
WER	Word Error Rate

NAMING CONVENTION

Words	Naming Convention
Bahasa Melayu words	In italics e.g. *ibu*, *abang*
English words	In quotes e.g. “tyrannosaurus”
Non-words	In quotes e.g. “idu”
Error types	In inverted commas e.g. ‘Reversals’, ‘substitutes vowel’
Alphabet or value or digit or syllable	In inverted commas e.g. ‘b’, ‘1’, ‘bu’
Emphasized word(s)	In bold face e.g. **dual-route model**
New terms	In italics e.g. *backpropagation*
Programming code or phonetic symbols (Worldbet) or file names or mathematical equation	In Courier New face e.g. `aku = A kh U`, `words.lexicon`

CHAPTER 1

INTRODUCTION

1.0 Introduction

Learning disabilities (LD) have gained serious attention from research communities in various fields – clinical, psychological, education, as well as computer science. LD is a condition where people have problems in acquiring skills essential in learning. Skills that are mostly affected are reading, writing, spelling, speaking, reasoning, and doing mathematics. The National Dissemination Center for Children with Disabilities, NICHCY (2004) revealed that 1:5 people are learning disabled. LD is a term generally used to refer to more specific types of learning problems such as Attention Deficit Disorder (ADD), Attention Deficit Hyperactivity Disorder (AHDD), and dyslexia.

Dyslexia is a type of LD that affects the individual's ability to learn basic literacy skills. The International Dyslexia Association (2006) defines it as a learning disability which is neurological in origin. Its characteristics are problems in accurate or fluent word recognition and also poor spelling and decoding abilities. These problems are strongly related to phonological deficits which are not clearly apparent in other cognitive abilities and from general classroom instructions. Other consequences include problems in reading comprehension and reduced reading experience that holds back the growth of vocabulary and background knowledge.

People who are dyslexics normally have average or high intelligence but found reading, spelling, and writing such overwhelmed tasks. Phonological deficit is acknowledged as being the cause of disability in reading and spelling and are predominant in children with dyslexia (Rainger, 2003). In Share and Stanovich study (as cited in Audiblox, 2000), the phonemic awareness, being a sub of phonological awareness is the major contributing factor of dyslexia. Phonological deficit refers to problems in decoding or deconstructing words into phonemes where it involves recognizing, sequencing, and producing letter sounds. For example, the second phoneme of these two words, *baca* and *beca*, differentiate them and gave them different meanings. The word *baca* means read in English and *beca* carries the meaning of a trishaw, a three-wheeled vehicle. Dyslexics have a lot of difficulties just to do that. Despite phonological decoding, other types of difficulties that usually affect dyslexics are as follow (Rainger, 2003):

- Visual processing
- Reading and comprehension
- Auditory processing
- Memory recall
- Structure and sequencing
- Planning and organizing

Even so, with the right help, children with dyslexia can be high achievers in education and in life (NICHCY, 2004). Help can be in many ways from the support from parents and teachers to specialists using devised programs to computer-aided programs and software. Technologies such as word processing, text-to-speech (TTS), and spell checker could be benefited from. Dyslexics can use these to facilitate reading and writing but these technologies could not train or teach them to grasp the required skills themselves. Automatic speech recognition (ASR)-based application is a potential tool for assisting children, especially dyslexics, in training them to learn the skills as noted in Russell et al. (1996).

ASR has been used to help university students in note taking and produce texts from lectures so that students can access the hardcopy and softcopy of the lectures being delivered (Bain, Basson, & Wald, 2002; Wald, 2004, 2005, 2006). The Liberated Learning Project (LLP) is a research on enhancing and implementing ASR technology to enable communication in university classrooms in assisting the disabled and learning disabled students including dyslexics (Wald, 2006). For dyslexic students, this technology is helpful in expressing written communication.

The LLP project is dedicated for helping dyslexic students already pursuing their tertiary education, despite the difficulties of which they struggle to cope with in earlier education. To ease the struggles, help should be provided as early as possible especially when dyslexic children enter their primary school years to make sure that they are not left too far behind. Projects such as the Colorado Literacy Tutor, CoLiT (http://www.colit.org/) with its component, the Reading Tutor Project is aiming at providing computer-aided reading instructions for children to enhance reading with collaborations with public schools (http://cslr.colorado.edu/beginweb/reading/reading.html). Another example of such project to improve reading amongst children is LISTEN's Reading Tutor (Banerjee, Beck, & Mostow, 2003a). These projects use ASR as the key technology to track reading and allow children to interact with the programs using speech (e.g. asking questions). Pronunciation accuracy is also provided for feedbacks.

ASR technology has a potential to enhance reading ability for normal children and it is also a potentially good tool for helping those with dyslexia in reading as reported by previous studies (Hagen, Pellom, Vuuren, & Cole, 2004; Nix, Fairweather, & Adams, 1998; Raskind & Higgins, 1999; Williams, Nix, & Fairweather, 2000). ASR is said to provide bimodal presentation of words (Higgins, 2004; Higgins & Raskind, 2000; Raskind & Higgins, 1999). Bimodal presentation means that a word is represented in a two-way manner where the word gets pronounced and written on screen simultaneously. This type of presentation gives dyslexics a multi-sensory experience, in which they can hear the words (spoken into the ASR) and see them being displayed on screen, which could increase their phoneme-grapheme associations (Raskind & Higgins, 1999; Snowling, 2000).

Exploring the potential of ASR to help dyslexics reading in Bahasa Melayu (BM) requires the ASR itself to be good enough to produce accurate recognition for that particular language. Hence, this study investigates how to model the ASR so that its performance is at its optimum when recognizing dyslexic children's reading isolated word in BM. Therefore, the effects dyslexia has on reading BM words to the children are considered as a contributor to the recognition ability of an ASR. Thus, with the fact that accuracy is indeed a major concern for any ASR system or application, how can the effects be incorporated into ASR for better performance?

To answer, the thesis is thus organized by presenting, in this chapter, the underlying challenges, objectives and scope to bind the research with the following chapters as the key towards understanding the whole study and what it has contributed to the field.

1.1 Research Taxonomy

The thesis highlights the ASR model proposal of dyslexic children's reading that includes the children's reading and spelling error patterns in BM. In reference to the Association for Computing Machinery (1998), the research taxonomy is as follows:

Knowledge area: Computer science and education

Field: Artificial Intelligence (I.2),

Computers and education (K.3).

Majoring: Natural Language Processing (I.2.7),

Computer uses in education (K.3.1).

Sub-major: Speech recognition and synthesis

Focus domain: Dyslexic children, reading, BM.

Dyslexic children are chosen as the focus domain for the study to obtain their reading and spelling patterns so that computer-based application, in this case an ASR, can be modeled with accordance to their readings. Dyslexic children when reading tend to produce a lot of reading mistakes that sometimes are not obvious in normal children. Their main symptom of reading errors is transpositions of similar shaped letters such as 'b' for 'd' or 'p' and vice versa, 'm' for 'w' and vice versa, and 'n' for 'u' and vice versa. Another reading error is made when they tend to substitute words with another, either semantically or not. Examples include substituting the word *pantai* with *tempat* for semantically related words and substituting *dan* with *dengan*.

This study focuses on proposing an ASR model for BM specifically for dyslexic children's reading and spelling patterns. Henceforth, the reading and spelling patterns is used interchangeably with reading patterns to reflect the errors they produced when reading isolated words. The error patterns must be in the context of BM errors produced by these children when reading aloud. The model is significant as it includes context-dependent pronunciation models of their readings together with the errors as means to improve recognition accuracy.

1.2 Motivation

As noted earlier, reading, writing, and spelling are three basic literacy skills that are crucial for children to master early. Late proficiency in these skills will have major impacts on the children's ability to grasp knowledge and will lead to low academic achievement. Enlightened by this fact, the Malaysian government through the Ministry of Education (MoE) has launched *Pelan Induk Pembangunan Pendidikan* (PIPP) under *Rancangan Malaysia ke-9* (Kementerian Pelajaran Malaysia, 2006). PIPP is aimed at enhancing the quality of education and improving the education system with six main agenda: build nation and people, develop human capital, strengthen national school, narrow education gaps, make teaching a prestigious profession, and make excellence a culture in educational institution (Kementerian Pelajaran Malaysia, 2006).

This study is motivated by the fourth agenda – to bridge the education gaps so that all schools and students in Malaysia will receive equal opportunities in education, regardless of locations, socioeconomic status, and the students' ability which includes the disabled and learning disabled students. The ministry through PIPP specifically stated that:

MoE offers education to the disabled students, including vision impaired, hearing impaired, and learning problems. Students categorized as having problems in learning are those with light cognitive problems, behavioral problems, autism, Down Syndrome, Attention Deficit Disorder (ADD), Attention Deficit hyperactivity Disorder (ADHD), and learning disabilities **especially** dyslexia. (Kementerian Pelajaran Malaysia, 2006, p. 95).

Students diagnosed with dyslexia will need special attention from parents, teachers, schools, as well as the ministry to ensure their active participations in education and that they are not left behind. Therefore, this study envision at helping dyslexic students to learn how to master literacy skills particularly in reading through the use of ASR-based application as assistive technology. This could only be realized by first having an ASR recognizer that is built for that task, i.e. with satisfying performance in terms of its ability to recognize BM words with high phonetically reading errors due to the children's difficulties. Inline with the government in helping disabled students in academic, the ministry also acknowledged that the government is willing to provide suitable ICT infrastructure among other helps as mentioned in PIPP section 7.22 of Chapter 7 (Kementerian Pelajaran Malaysia, 2006).

1.3 Problem Statement – Challenges at Hand

With advancing technology in educational environment, computer-based applications are being used as teaching aids to boost the children's interest and motivation towards reading. An exhaustive list of examples of software for dyslexic children to learn how to read is available at http://www.dyslexia-magazine.com. However, most of the teaching software are equipped with only text-to-speech (TTS) to provide help only when requested by dyslexic children. Quite often, these children are not aware of the errors made when reading aloud and thus do not ask for help. For dyslexic children to learn to read efficiently, immediate intervention by teachers or reading facilitators is required during the actual reading. Immediate intervention means that the teacher or reading facilitator should immediately interrupt the children's reading when an error is encountered. This is to ensure that learning process continues and that the children are aware of the correct pronunciation of a particular word.

To enable and automate such intervention in computer-based applications, ASR technology is therefore being used. This technology plays a significant role in recognizing what have been read by the children so that any mistake made during reading can be detected and hence intervene the process. Without ASR technology, this reading tracking mechanism could not have been materialized. The use of ASR technology in this area imposes challenges in two aspects:

- To date there is no dyslexic children's reading model for BM that has been reported. The current models are all of English language reading models based on cognitive processes that do not list the effects dyslexia has upon the processes that yield such reading errors.
- There is no ASR model to support dyslexic children's reading in BM where the error patterns of reading and spelling are highly phonetically similar. This is a challenge for ASR since recognition of phonetically similar vocabulary is considered to be rather difficult towards achieving high accuracy and therefore lower word error rate (WER).

The first aspect concerns with the error patterns and their relation to reading models. Current reading models, although established the cognitive links involved in reading, do not include the effects of broken links affected by dyslexia, which contribute to reading and spelling error patterns. It is of significant that the patterns are included to highlight the possibilities of error patterns which failure of the broken links contributed to. Therefore, this condition has raised a few questions such as "what are the reading and spelling patterns for dyslexic children reading in BM?", "is there any specific vocabulary for dyslexic children to learn to read?", "can a reading model be developed to model dyslexic children reading in BM?" and "can the reading model be designed and developed by enhancing existing model(s)?".

The second aspect is in terms of speech recognition for children's speech. Current successful applications of ASR are based on adults' speech that differs significantly from children's speech (Das, Nix, & Picheny, 1998; D'Arcy, Wong, & Russell, 2004). Despite the significant differences in speech produced by children due to differences in pitch, frequency, and vocal track length among others, children often produced softer speech when asked to read aloud. Although sharing the same properties of normal children in terms of the production of speech, dyslexic children also have slower articulation of speech as demonstrated by Kasselimis, Margarity, and Vlachos (2006). They too often have problems with the on-set and rime level as well as syllabic level (Goswami, 2002).

Another issue for consideration is the BM corpus. Most available ASR tools are in English. There are ASR applications in BM but they do not cater for dyslexic children especially one that includes the BM error patterns of reading and spelling. For Malaysian dyslexic children, they too need to learn to read in BM as it is the official language of the country. So far no work on ASR that includes reading and spelling errors of BM for dyslexic children has been reported. To enable speech recognition in BM that includes reading errors or miscues requires an ASR model that is built according to the reading and spelling error patterns of dyslexic children. Again, this limitation has risen questions such as "can an ASR model be modeled based on the previous reading model?" If so, can the model be designed and developed based on existing ASR model(s)? In addition, can the model be evaluated and what are the criteria to take into consideration when evaluated?

1.4 Research Objective

The main objective of this study is to propose an ASR model that is designed for dyslexic children's reading model in BM. The aim is for the ASR to be able to recognize dyslexic children's reading and detect miscues to envision a tracking mechanism in BM. The study focuses on word recognition, which is the first level of reading. Given the main objective, the specific objectives of this study are as follows:

a. To collect selected vocabulary within the BM level one primary school syllabus.
b. To recognize dyslexic children's reading mistakes and classify their reading error patterns.
c. To improve dyslexic children's reading model in BM based on findings in (b).
d. To extend ASR model for that particular reading model as in (c).
e. To develop an ASR recognizer based on (d) for testing and evaluation.
f. To evaluate the proposed ASR recognizer in terms of accuracy measured in word error rate (WER), miscue detection rate (MDR), and false alarm rate (FAR).

For specific objective (a), the BM vocabulary suitable for children is collected based on simple words that are commonly used in primary school syllabus. The vocabulary need to match the dyslexic children's reading level (level one or *tahap satu*). It is of significance to note that due to dyslexia, these children might portray poorer reading and so the words need to be carefully selected. The selected words represent 23 syllable patterns, as specified in the syllabus, that could reflect the reading abilities of the dyslexic children in order to obtain their reading and spelling errors.

As for objective (b), the reading and spelling errors or miscues made by dyslexic children are recognized and classified into suitable categories. Noteworthy, since this study works on read words rather than spoken attributes, the pronunciation types are identified based on the children **reading aloud**, not while they are speaking. The error categories serve as important elements for modeling pronunciations of the selected words where only the three most frequent error categories are modeled for generalization. However in this case, there are two error categories sharing similar percentile and thus the total of the most frequent error categories is four.

To achieve objective (c), the frequent error categories are introduced into an existing reading model (word recognition level) as an improvement of the model. The introduction of the dyslexic children's frequent error patterns into the reading model provides a deeper understanding on how dyslexia could affect reading and consequently what are the effects, i.e. the frequent reading error patterns.

For objective (d), the ASR model is specifically designed by incorporating reading error patterns obtained. For each particular pattern, a pronunciation model is constructed for each of the words involved. The pronunciation models serve as ASR lexical model for which training is later performed. Prior to training, an ASR model is chosen for further improvement. The improvement made to the model concerns with the lexical model and how it is constructed and represented so that recognition accuracy is not compromised.

For objective (e), with the ASR model proposed a recognizer is developed to incorporate the findings identified earlier. The ASR recognizer is trained to recognize single, isolated BM words of selection, which are modeled into the lexical model according to the ASR model proposed.

Finally, achieving objective (f) requires testing and evaluation of the ASR model in terms of its recognition accuracy. The recognition accuracy is measured using WER for its performance in terms of recognizing the words (the lower the better). Since the study concerns **reading**, it is of significance to measure its ability to recognize miscue and false alarm, measured in MDR and FAR respectively.

1.5 The Scope

The domain of this study concerns Malaysian children with dyslexia whose difficulties in reading are still at the very basic level that is word recognition. The dyslexic children participated in this study are of primary school age between 7 to 14 years old[1] who posses similar reading level as identified and suggested by their teachers. Most ASR applications use readings from children ranging from as early as 5 years old to 18 years old (Aist et al., 1998; Fairweather, Nix, Oblinger, Adams, & Carla, n.d.; Hagen, Pellom, & Cole, 2003; Higgins & Raskind, 2000; Nix et al., 1998; Raskind & Higgins, 1999;

[1] Dyslexic children are allowed a maximum of two years extension of primary schooling in the Malaysian education system. This enables any dyslexic child who is not ready for secondary school to stay in the primary school until the age of 14. In addition, dyslexic children are also required to follow the main stream syllabus and examination (where a *reader* is provided), and only attend to special dyslexia classes taught by special education teachers for subjects such as BM, English, Mathematics, and Science.

Russell et al., 1996; Shobaki, Hosom, & Cole, 2000). For this study, a total of ten dyslexic children are involved, 8 males and 2 females. The number of children participated in the aforementioned studies ranges from as small as five to a few thousands. Thus, having 10 dyslexic children participated in this research is considered representative according to various researches (Russell et al., 1996; Nor Hasbiah, 2007). In fact, it is adequate for achieving the desired results, aiming at the ASR model that could generate significant recognition accuracy.

The participants were recruited from two primary schools – Sekolah Kebangsaan Jalan Datuk Kumbar, Alor Setar, Kedah and Sekolah Kebangsaan Taman Tun Dr Ismail (2), Kuala Lumpur. These schools are among primary schools that provide special dyslexia classes, a pioneered program launched and fully supported by the MoE.

In line with the focus of the MoE of which to facilitate learning for dyslexic children in BM, the language of concerns is BM at word recognition level. Word recognition level is the basic level of reading (Ellis, 1993), often involves early readers. Early readers are usually introduced to simple BM words in learning to read. Thus, since this study concerns dyslexic children whose reading level matches or similar to that of a six to eight year-olds, word recognition level is thus most suitable.

To learn to read BM words, the controlled vocabulary selected for this study therefore consists of words included in the BM syllabus specified by the ministry. The words appear in Year 1, 2, and 3 text books as well as *Buku Panduan Pelaksanaan Program Pemulihan Khas (Masalah Penguasaan 3M)*. The vocabulary is chosen based on suggestion made by BM teachers in the aforementioned schools. The justification is that the vocabulary should contain common words within level one syllabus as suggested by various studies (Russell et al., 1996; Baumer, 1998). The vocabulary consists of BM words that correspond to 23 syllable patterns as listed in the *Buku Panduan Pelaksanaan Program Pemulihan Khas (Masalah Penguasaan 3M)* provided by the ministry. These patterns are constructed by consonant-vowel pair (C-V pair), e.g. the word *itu*, which is constructed based on V+CV syllable pattern and *baca* that falls under CV+CV syllable pattern. The syllable patterns listed include also prefix and postfix forms of C-V pairs, e.g. prefix 'me-' of CV pair and 'pen-' of CVC type. The vocabulary is used as stimuli for the reading recording sessions where a participant is prompted a word and asked to read that word out loud. The readings are recorded to capture the speech signal for further analysis.

The study focuses on proposing a reading and spelling model for dyslexic children reading aloud isolated words in BM and developing an ASR model based on the reading model. In future this is envisioned to be of help for dyslexic children to grasp spelling and reading skills using an ASR-based application that incorporates the ASR model. For the purpose of this study, only the words that fall under the most frequent error patterns were used.

As for the ASR context, the speech files are limited to 289 files of 16 kHz (16000 samples per second). The speech files are automatically divided into three datasets – 188 for training, 53 for development, and 48 for testing the ASR recognizer. Another 100 speech files of the same specification are used for the purpose of miscue detection evaluation. Since the recordings were performed in computer laboratory environment, it can be observed that there is substantial noise present which affects the quality of the sampled speech. Therefore, the speech files involved are bounded by the most frequent error patterns and the quality of the sampled speech.

1.6 Contribution of Research

Achieving the objectives lead to significant contributions to knowledge, special education teachers and reading facilitators, parents, researchers, software developers, and the MoE. These contributions, not only significant but also open a wide opportunity for future research in both area of dyslexia and reading in BM as well as in ASR technology. The research significant contributions are:

- The reading model of dyslexic children, which modeled the four most frequent error patterns surfaced when reading isolated BM words as the effect or outcome of broken links of cognitive processes in word recognition.
- The extension of a hybrid Hidden Markov Model and Artificial Neural Network or HMM/ANN-based ASR model that suggests that the lexical model and the language model are emphasized to highlight their importance towards recognition accuracy. In addition, the phonetic refinement rule and the most frequent errors are incorporated into the lexical model for optimum performance as presented in Chapter 4 and 5.
- An ASR recognizer as a working model that demonstrates its ability to recognize readings of selected words in BM with optimum rates. The recognizer can be extended to incorporate suitable interface for a full working model.

The research's benefits of proposing these two models are threefold. One, the proposed reading model of dyslexic children reading in BM can serve as a basic model towards the development of knowledge in education as well as in ASR technology. This model together with the reading error patterns surfaced in this study suggest to a teaching method that is tuned towards BM spelling system with further enhancement in related area such as education as well as cognitive-based and behavioral studies. Dyslexic children could benefit from the teaching method that is targeting to correct BM reading, which differs from English. Special education teachers and reading facilitators can use the reading model to guide their teaching. They can also use the error patterns as targets of remediation where the most significant patterns are addressed and corrected to enhance reading performance of the children and at the same time increasing their confident towards reading. The same also contributes to parents whose children are dyslexics.

Two, the extended ASR model serves as a fundamental building block towards future development of a BM reading tracking mechanism for realizing an automated reading instruction. Should this be materialized in future, dyslexic children could gain positive benefits from using such application that is specifically designed for them. In addition, the ASR model suggests that it could be extended to incorporate more error patterns for further enhancement. Researchers and software developers can work hand-in-hand to build an ASR-based application using the ASR model as basic building block for the reading tracking mechanism in a computer-aided reading instruction. With the help of both the models and the findings of the patterns surfaced, dyslexic children can benefit from a more suitable, BM-tuned teaching method and an enhanced ASR-based application to remediate their difficulties in BM.

Finally, the MoE is able to bring the education gap closer for normal and disabled and learning disabled children as addressed in the 4th agenda of PIPP. Both the models can be further extended to realizing the mission and vision of PIPP where an ASR-based application to teach reading to the children can be further researched and extended. Even the reading model alone, is able to give information to the R&D department of the ministry. It could be used to devise a special method to tackle BM spelling and reading to improve their performance in reading and positively increase their participation in academic.

1.7 Overall Thesis Summary

With the objectives, scope, and contributions identified, the study embarks to review related literature and suitable methodology before continuing to the development of an ASR recognizer as means for further evaluation. The thesis is thus a systematic report of the research conducted, which is focusing on proposing an ASR model that is based on dyslexic children's reading and spelling error patterns in BM. Next subsection summarizes the study and the following subsection presents the thesis organization.

1.7.1 Research Recapitulated

With regards to the objectives as specified in Section 1.4, the research is conducted to achieve those objectives. Table 1.1 describes, in summary, research problems of the study that probe research questions and objectives. Methods to achieve all objectives and expected deliverables are also outlined.

Table 1.1. Research problem and research questions addressed in this study together with summary of methods and expected deliverables.

Research Problem	**Research Questions**	**Objectives**	**Methods**	**Expected Deliverables**
There is no dyslexic children (DC)'s reading model for BM that has been reported	1) Is there any specific vocabulary for DC to read? 2) What are the reading error patterns for dyslexic children reading in BM?	1) To collect BM vocabulary, and construct pronunciation models. 2) To identify dyslexic children's reading error patterns.	Literature; Informal interviews with special education teachers and dyslexia facilitators; Ethnography: observation and recording;	Fundamental theory and elements of reading & spelling error patterns
	3) Can a reading model be developed to model DC reading in BM? 4) Can the model be designed and developed using existing model(s)?	3) To model DC reading error patterns in BM	Informal interviews with special education teachers and dyslexia facilitators; Literature; Data analysis	DC reading model in BM
There is no ASR model to support DC reading in BM which is based on the previous reading model.	5) Can an ASR model be modeled based on the previous reading model? 6) Can the model be developed based on existing ASR models?	4) To model an ASR model for that particular reading model.	Literature; Results from data analysis.	ASR model to support reading in BM

7) Can the proposed model be evaluated?	5) To develop a recognizer based on the ASR model proposed.	Rapid Application Development	ASR recognizer, evaluation results, list of elements to further improve the model
8) What are the criteria for evaluation?	6) To evaluate the proposed ASR model.	Evaluation – WER and MDR, FAR.	

1.7.2 Thesis Organization

This thesis is organized in line with the specified objectives. Each of the steps taken to achieve all the objectives is discussed in the consequent chapters. To meet the purpose, the thesis is thus organized in six chapters.

Chapter 2 presents extensive review of literature composed of both domains – dyslexia and ASR technology. The review of literature for dyslexia outlines the background of this condition and its effects towards reading development among children including the theories behind it. This chapter also links dyslexia and reading to ASR technology, discussing the ASR types, methods, and models.

A comprehensive and in depth coverage of the approach and methods the research employs is described in Chapter 3. The methods are presented in a step-by-step discussion of each process involved aiming at achieving the objectives as outlined in section 1.4. In addition, the research findings in terms of the most frequent spelling and reading error patterns that structure the reading model are also presented as the basis towards the development of an ASR recognizer that reflects the ASR model proposed.

To further discuss the development in details, Chapter 4 reports on the development of the recognizer that employs the ASR model proposed. The development is discussed and addressed as means for further evaluation of the proposed model.

Chapter 5 presents the evaluation results and discusses an extensive analysis on the evaluation results. The metrics of concerns are WER and MDR-FAR. WER is a widely used metric to measure the ASR recognition accuracy based on the percentage of errors. The lower the percentage is the better. It is directly linked to recognition rate that gives the percentage of correct recognition by subtracting WER from a total 100%. MDR is used to measure the accuracy of the recognizer in detecting reading miscues. Therefore, higher rate implies better performance with regards to the FAR performance.

The thesis is then concluded in Chapter 6 where the achieved objectives are re-emphasized to highlight the methods and deliverables. This chapter also discusses the limitations and challenges in important aspects of the study such as the model development of both the reading model and the ASR model. Open gaps are also addressed for future research directions in related fields.

CHAPTER 2

LITERATURE REVIEW

2.0 Introduction

Reading performance of dyslexic children are below their age when compared to their peers. ASR technology is seen to facilitate the process of reading development as it gives the multi-sensory experience to stimulate their brains. Studies looking into ASR technology in reading development amongst dyslexic and non-dyslexic children have found that this technology does play a significant role in enhancing reading performance for these children (Higgins, 2004; Higgins & Raskind, 2000; Raskind & Higgins, 1999; Russell et al., 1996; Williams et al., 2000). This technology is found to offer such effect as it can remediate the problems that concerns with phonological representation (Higgins & Raskind, 2000; Raskind & Higgins, 1999). ASR applications can be used to train dyslexic children in reading through series of practice where children are required to read into the computer (Higgins, 2004; Higgins & Raskind, 2000; Mostow, Roth, Hauptmann, & Kane, 1994; Nix et al.,1998; Raskind & Higgins, 1999; Williams et al., 2000). What makes the remedial effect is the multi-sensory experience offered by using this technology.

The multi-sensory experience is created through a bimodal presentation of words – pronunciation (hear) and spelling (see) as the computer display the word on screen – serves as the key feature for the remediation. Correct pronunciation will yield correct spelling and thus gives them the opportunity to learn to spell at the same time. Furthermore, it is also noted that the correction procedure does play a significant role in reading development among these children. The correction procedure, should the children make any mistake, requires the children to analyze the spelling of a list of similarly spelled words (presented in a choice box) and choose the correct spelling. Although Raskind and Higgins (1999) claim that this strategy to be a remedy for reading, it still requires moderate reading comprehension towards selecting the correct spelling of words in the correct context (University of Washington, 2002). It seems that this requirement alone contributes to unsuitable use of ASR technology to aid reading development for children with severe dyslexia. Even so, the results of previous studies have shown that a significant increase has taken effect in dyslexic children while using ASR applications to enhance their reading performance (Hagen, et al., 2003; Hagen et al., 2004; Higgins & Raskind, 2000; Nix et al., 1998; Raskind & Higgins, 1999; Raskind & Shaw, 1999; Russell et al., 1996; Williams et al., 2000).

ASR technology is in fact a promising alternative to conventional methods of teaching dyslexic children to read. Other than it offers the most essential key to learning for dyslexic children ASR technology also enables interaction between the children and computers. Studies have found that children are more excited and less intimidated to try to utter unfamiliar or difficult word plus, using computers is just attractive enough compared to conventional teaching (Lerner, 1997; Olson & Wise, 1992; Russell et al., 1996). Furthermore, a study demonstrated that using computers to reduce reading difficulties manages to improve children's performance in terms of their phonological awareness, word recognition, and letter naming skills (Mioduser, Tur-Kaspa, & Leitner, 2000).

This chapter is organized as follows: Section 2.1 brings forward the theory that underlies this study. Section 2.2 delineates the vocabulary and suitable age group that is appropriate for the words in the vocabulary. Section 2.3 reviews the reading models that captures the modules or processes involved in word recognition and reading aloud. Next, Section 2.4 reviews types and methods of ASR technology that could be employed in the study. Section 2.5 discusses how accurate an ASR is and specifies the percentage of word error that is frequently used as metric for the purpose of evaluation, as well as the percentage for miscue detection. And finally, Section 2.6 summarizes the chapter.

2.1 Using ASR Technology to Facilitate Reading: Theory-based Perspective

Fluent reading involves two important elements – word recognition and reading comprehension. Word recognition is vital to enable automaticity in reading process in that it allows us to be able to recognize words and decode unfamiliar words. Reading comprehension on the other hand, is to understand the meaning of what we have read. In order to understand the meaning, one must be able to recognize the word. That is why young children are taught to master word recognition skills with little emphasize on comprehension before going for level 2 (standard 4 – 6 in primary schools) which focuses more on the latter skills. Therefore, agreeing with the idea that reading should be taught at early age, this study only focuses on the very first step to fluent and effective reading – word recognition skills. To master word recognition skills, one should be able to match graphemes (letters) with their appropriate phonemes (sounds), which directly lead us to phonological processes.

Unfortunately for dyslexics, phonological processes have been the most dominant factor for failure in reading for many studies have recorded that dyslexic children's performance were lower than their normal peers in this area (Bradley & Bryant, 1983; Snowling, 2000; Stanovich & Seigel, 1994; Vellutino, Scanlon, & Spearing, 1995; Vellutino et al. 1996; Vellutino, Fletcher, Snowling, & Scanlon, 2004). Hence, phonological processes are no doubt the critical factor for fluent and efficient reading. In fact, many have agreed that *phonological deficit theory*[2] is the universal theory that underlies the relationship between dyslexia and reading (Frost, 2001; Lundberg, 1995; Shaywitz, 1996; Snowling, 2000; Wolf, 1999; Ziegler, 2006).

Based on the phonological deficit theory, another theory has been developed that introduce another factor which deficit in this skill may lead to reading difficulties – the *double deficits theory* (Wolf, 1999; Wolf, Bowers, & Biddle, 2000). This theory is not deviating from the universal theory of dyslexia but rather introduces naming speed as a co-factor and together, both phonological awareness and naming speed, contribute to reading skill development. Reading problem is regarded as severe if phonology and naming speed are both impaired.

Unlike the phonological deficit theory, which regards naming speed as a sub of phonological processes, the double deficits theory suggests it otherwise (Wolf, 1999; Wolf et al., 2000). Though this theory sounds suggestive, it is lacking evidence to support the hypothesis that dyslexia is attributed to phonological deficits or/and naming speed deficits. For example, Manis, Custodio, and Szeszulski (2000) have found that naming speed deficit does not affect word recognition skills and in fact, children with

---

[2] This theory is also referred to as phonological-core deficit theory or phonological-based theory. The terms are used interchangeably to refer to the same theory.

only naming speed deficit has at least average word recognition skills. Therefore, the role of speed naming in reading skills can be questioned.

It is argued that phonological deficit theory stands stronger than the double deficit theory and thus, the use of multi-sensory approach is critical to compensate phonological deficit and further help remediate reading process. For dyslexics, multi-sensory method is necessary so they can learn to read by using other senses. This is where the use of computers and ASR technology come in handy. Although there are conventional methods that offer multi-sensory experience too, dyslexic children have shown significant improvement in reading performance when using computers as demonstrated by various studies (Conn & McTear, 2000; Dwyer, 2000; Higgins, 2004; Higgins & Raskind, 2000; Lundberg, 1995; Olofsson, 1992; Olson & Wise, 1992; Raskind & Higgins, 1999; Raskind & Shaw, 1999).

Noteworthy, many studies looking into the effectiveness of using computer for remediation focus on synthetic speech technology (Lundberg 1995; Lundberg & Olofsson, 1993; Olofsson, 1992; Olson & Wise, 1992). Although they manage to positively increase reading performance for the children, remediation using computers and synthetic speech has some limitations. What if the users (in this case the dyslexic children), have no idea that what they have read is wrong? It might be the case that they just simply guess the word or replace it with another word assuming they were correct. Then, most probably they might not click the 'help' button provided to get the computer to 'say' the correct word. This is where ASR technology fits in perfectly – to track their reading such as words, positions in text, and reading error as well as to provide assistance when necessary in order to trigger useful feedback so that the learning process continues. Studies have shown that such automatic reading tutor can bring not only motivation and encouragement but also improvement towards reading ability in children (Fairweather et al., n.d.; Mostow et al., 1994; Mostow & Beck, 2003; Nix et al., 1998; Williams et al., 2000).

Evidently, ASR is a potential technology to be used to help dyslexic children reading in BM. Thus, in line with the main objective and sub-objectives as defined in the previous chapter, the following sections provide discussions to support each objective: a) to collect selected vocabulary within the BM level one primary school syllabus; b) to recognize dyslexic children's reading mistakes and classify their reading and spelling error patterns; c) to model dyslexic children's reading in BM; d) to model an ASR model for that particular reading model; e) to develop an ASR recognizer for testing and evaluation; f) to evaluate the proposed ASR recognizer in terms of accuracy measured in word error rate (WER) and miscue detection rate (MDR)-false alarm rate (FAR).

2.2 The Vocabulary and Age-Matched Dyslexic Children

The must have element of identifying the spelling and reading pattern is the vocabulary. Here, the choice of words is of concern. Automatic reading tutors suggest using words from story books or reading materials for children (Aist et al., 1998; Fairweather et al., n.d.; Hagen et al., 2003; Mostow et al., 1994; Nix et al. 1998). These words are being used mainly to bring out the positive effect for the children towards reading such as enjoyment, encouragement, smoother reading, and comprehension (Baumer 1998; Lerner, 1997; Williams et al., 2000). However, the interest of this study is to collect vocabulary that best represents BM spelling patterns to enable improvement in word recognition for BM reading. Common words are best for such purpose as suggested in discussions with special education teachers that teach BM to dyslexic children. This opinion is supported by Baumer (1998), who suggest parents to use some vocabulary from the 'most used words' list taken from Dolch's 202 most common words in English for word recognition plus 30 more frequently used words. Like Baumer, Shobaki et al. (2000) also use the most common English biphones to cater for younger children's vocabulary in building the OGI KIDS corpus. Another work is reported to use 1000 word Primary School Reading vocabulary (Russell et al., 1996). BM corpus collection is also available that contains words taken from newspapers, literary books, academic writings, and user guides (Zaharin, 2000). This corpus is an effort from a machine translation research team called Unit Terjemahan Melalui Komputer (UTMK) of Universiti Sains Malaysia. This work provide written corpus as reported by Ranaivo-Malancon (2005).

It is argued that, in the case of having an ASR application for children reading in BM, the vocabulary chosen for the study covers common words under all consonant-vowel (CV) syllable patterns as presented in the syllabus. That is why the words are taken from *Bahasa Malaysia* Year 1, 2, and 3 text books plus some words from *Buku Panduan Pelaksanaan Program Pemulihan Khas: Masalah Penguasaan 3M* (Jabatan Pendidikan Khas, 1999). For example, the word *ibu* has two syllables – 'i' and 'bu' where the syllable pattern falls under the V+CV group. Another example is *kumpulan* which has three syllables 'kum', 'pu', and 'lan' in the CVC+CV+CVC group. The CV patterns differ from that of English (see Appendix A for a list of syllable patterns in words in BM). The reason for using common words is to provide easier reading in order to establish stronger word recognition skill as well as exposing the children with the variety of syllable patterns in BM. To read in BM fluently, one has to be able to know how to deal with these patterns for correct pronunciations. In this case, common words refer to words that the children have and shall encounter and learn in their classes, i.e. words within the syllabus.

Once the vocabulary is constructed, the next objective concerns with obtaining the reading error patterns. Thus, the following section reviews previous works focusing on the errors related to reading and spelling of dyslexic children.

2.3 Reading and Spelling Error Patterns

It is worth to look at the spelling error patterns of dyslexic children since reading and spelling are closely interrelated skills (Reid, 2003; Sawyer, Wade, & Kim, 1999). Spelling is crucial towards fluent and correct reading. Spelling error patterns are also important in identifying errors made when a child tries to read aloud a word (word level reading) since it involves either direct word recognition (often with very familiar words) or spelling for other words. However, according to Sawyer et al. (1999) little is known about the spelling patterns of dyslexic children and thus, they carried out a study to identify the spelling error patterns of dyslexic children which resulted in patterns as illustrated in Table 2.1. The categories are adopted from Moat (1995) as cited by Sawyer et al.

Table 2.1. Phonological error types and their definitions and example words adapted from Sawyer et al. (1999).

Categories of Patterns	**Definitions**
Adds consonants	Adding a consonant ("whent" for "went"; "wripe" for "ripe")
Adds final *e*	Adding final *e* ("fede" for "fed")
Incorrect sequence	Inappropriate sequence of letters ("help" for "help")
Liquids (add/delete/substitute – *l* or *r*)	Addition, deletion, or substitute *l* or *r* ("sed" for "slid")
Nasals (add/delete/substitute – *m* or *n*)	Addition, deletion, or replace *m* or *n*. ("stemp" for "steep"; "bop" for "bump")
Omits final *e*	Omission of final 'e' ("cap" for "cape")
Omits vowel	Omission of a vowel ("cost" for "coast")
Omitted consonants (excludes *m, n, l ,r*)	Errors of consonant omissions (*sep* for *ship*; "tip" for "trip")
Reversals	Reversing the form of a letter ('b' for 'd' or 'p')
Substitutes consonants	Substitution of one consonant with another ("jem" for "drum"; "fid" for "fit")
Substitutes nasals for liquid	Substitution of a nasal for a liquid ("nap" for "lap")
Substitutes vowel	Substitution of one vowel with another ("desh" for "dish")
Substitutes word	Substitution of one word for another ("it" for "chop")

In this study, they have found that the most errors made is vowel substitutions (805 errors or 39.75%) whereas consonant substitutions and consonant omissions collectively came second. Hence, the identified categories or spelling error patterns serve as the base of this research. Few categories might be dropped and few more are added depending on the findings. As expected that a few changes are necessary since we are dealing with different language with different orthographic system.

Due to the phonological deficit, dyslexic children have certainly somewhat different reading patterns than that of normal peers. The one thing that distinguishes them from other garden-variety poor readers (of same chronological age) is that despite adequate exposure to the language, they read poorly at least two years below their chronological age. The dyslexic children's reading errors in general are: repetitions, additions, deletions, substitutions, reversals of letters/numbers, transpositions. It is important to note that these errors are of *sentence reading* or *sentence level* reading where the children is reading a passage or text. Since the focus of this study is of word recognition, which is the first level of learning to read, the study thus focuses on *word-level* reading. Word level reading is when a child reads one single word, often when the child is an early reader or a beginner.

The reading and spelling error patterns serve as the elements for modeling a reading model specifically for dyslexic children reading isolated words in BM. The next section reviews various reading models and selects a suitable model to be adapted for improvement.

2.4 Dyslexic Children's Reading Model

Reading involves certain cognitive processes to be activated when written word is presented. Reading does require different skills at different cognitive processes (Vellutino et al., 2004). Fortunately for most of us, these skills are automatically and unconsciously developed as we grow older and hence, reading skills seem to be grasped and mastered automatically. To illustrate the processes involved, a few reading models are discussed. The reading models are all of normal readers' cognitive processes towards reading or processing a word. Dyslexics do share the same processes in reading but unfortunately for them, some of the routes of a process to another are impaired.

The first reading model suggests that reading involves two separate routes, lexical and non-lexical, and is called the *dual-route model* (Coltheart, Curtis, Atkins, & Haller, 1993; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). This model considers three paths of processing a word - the lexical semantic and lexical non-semantic paths in the lexical route and the non-lexical route. Figure 2.1 demonstrates the routes and the processes involved. The lexical route is denoted by the solid arrow whilst the dotted arrow represents the non-lexical route to reading aloud. The lexical semantic path function is to process familiar words, including irregular words by triggering and activating the semantic system whereas the lexical non-semantic path does not. On the other hand, the non-lexical route's function is to process regular and non-words that do not conform to the typical grapheme-phoneme system (Castles, 2006). In other words, according to this model the non-lexical route is used to recognize words with inconsistent grapheme-phoneme representation such as "chaos" and "chart".

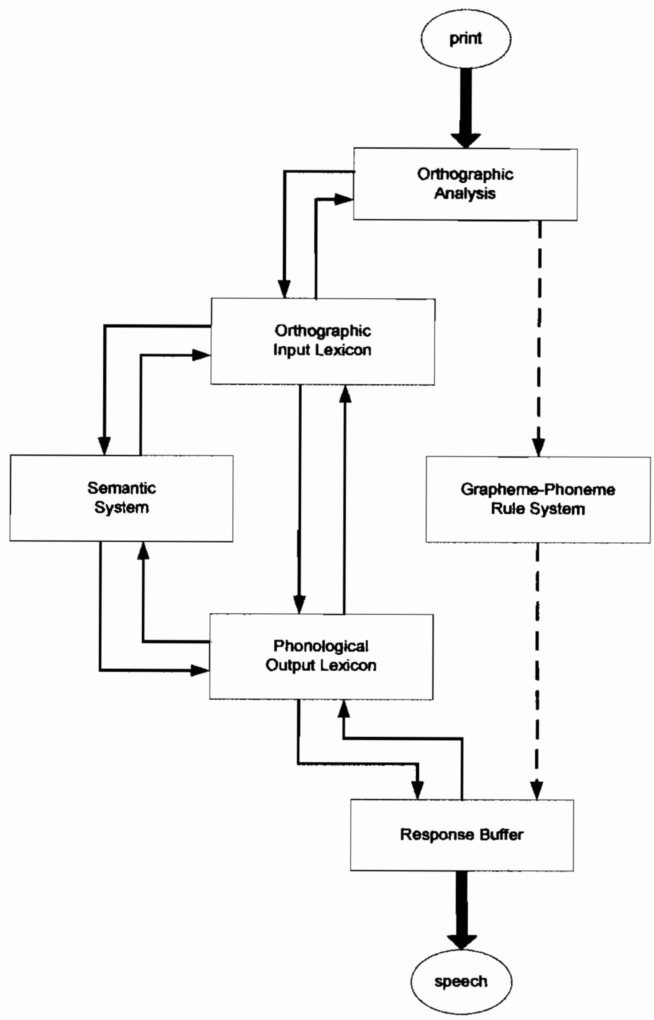


Figure 2.1. Basic architecture of the dual-route model adopted from Coltheart et al. (2001).

In contrast with the dual-route model, Seidenberg and McClelland (1989) claimed that their model, a computer simulated model called parallel distributed processing (PDP) or the *connectionist model*, can perform the same tasks without the need to have more than one route. Their assumption is that "...there is a single, uniform procedure for computing phonological representation from an orthographic representation that is applicable to irregular words and non-words as well as regular words." Figure 2.2 illustrates how this simulation model works.

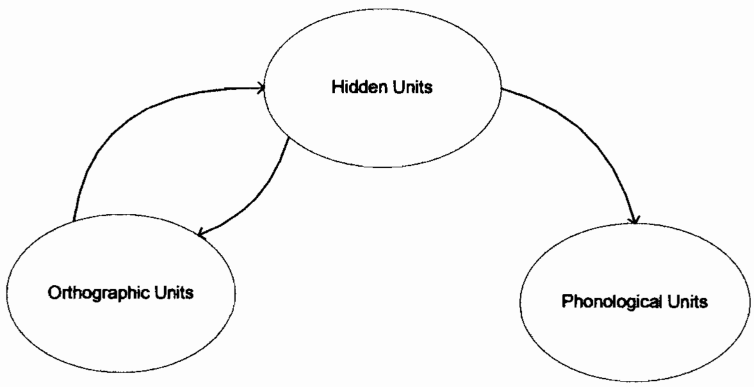


Figure 2.2. The structure of the Seidenberg-McClelland's computer simulated model of reading, adopted from Seidenberg and McClelland (1989).

Basically, this model uses *backpropagation* technique to train the units involved in transforming a word (orthographic units) to its correspondence pronunciation (phonological units). Seidenberg and McClelland (1989) implemented 400 orthographic units connected to all 200 hidden units, which are also connected to all 460 phonological units. The weights of all connections are random and consequent adjustment of weights is made as the network is trained.

The two models discussed have been much of a debate. The issue circles that whether a reading model should have one or more routes for processing written text to produce its corresponding pronunciation and whether one model is superior to the other. Such interesting debate can be found for example in Seidenberg and McClelland (1989), Coltheart et al. (1993), and Coltheart et al. (2001).

Note that however, the intention of this study thus far is not to argue whether a reading model should have one, two or more routes of processing a word. Instead, the aim is to see which model fits the purpose of dyslexic children's reading in BM bounded by the scope of this study. In this case, there are no non-words involved. So, if based on the dual-route model the non-lexical route is not used in the reading process. On the account of the connectionist model, no explicit representation of processes or modules needed in the reading process is presented. Instead, this model depends on the backpropagation to derive such flow. Therefore, another general but comprehensive reading model is considered.

Similar to that of the dual-route model, Ellis (1993) describes a simple model of reading, eliminating the non-lexical, grapheme-phoneme system rule as depicted in Figure 2.3. According to Ellis, this model conforms to and emphasizes on the broad agreement of reading models and avoids the controversial issues concerning the various models. Hence, this model is chosen along with the notion that it fits the study well where only word recognition level is considered. This model is represented by five important modules keeping intact the lexical route of reading, thus making it suitable for this study.

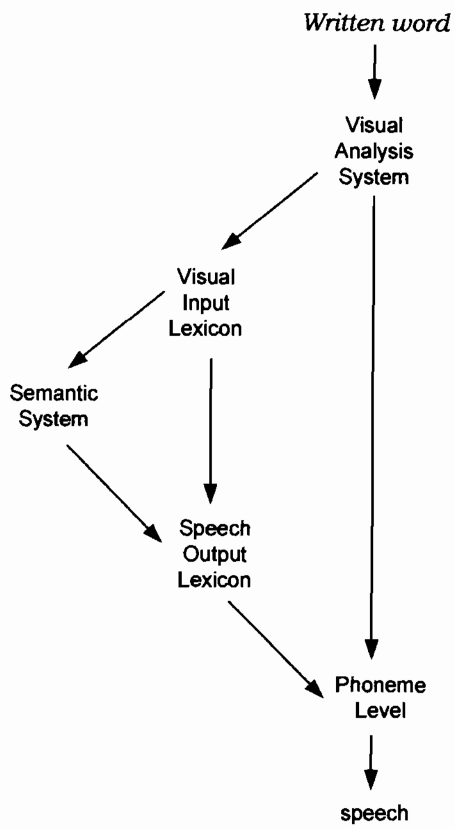


Figure 2.3. A model of cognitive processes involved in word recognition of a single word, adopted from Ellis (1993).

The modules are visual analysis system, visual input lexicon, semantic system, speech output lexicon, and phoneme level. Each of these modules plays different roles towards processing a word. The modules and their functions are summarized as follows:

- **Visual Analysis System –** This system performs two major functions. The first one is concerning the ability to identify letters regardless of the different shapes that they can be in (e.g. *satu*, satu, and satu). The other function is to note the position of letters in a word since the order of each letter in a word distinguishes that word from another (e.g. *tanya* and *nyata*).

- **Visual Input Lexicon** – Visual input lexicon module can be regarded as a mental word store or memory. Its function is to identify whether or not a presented word is familiar. This module can also be denoted as the word recognition units since it provides the path towards the meaning of each word and their correct pronunciation. For example, when presented with a familiar word such as *ibu*, visual input lexicon module is able to recognize this word but not the non-word such as "ibe" or "idu".
- **Semantic System** – This module contains the representation of all knowledge of meanings of familiar words. Consider the previous example, when presented with the familiar word *ibu*, the visual input lexicon module could retrieve the meaning of this word (mother) but fail to obtain the meaning of the other two non-words in BM, 'ibe' and 'idu' because their meanings do not occur anywhere in the semantic system.
- **Speech Output Lexicon** – The function of speech output lexicon module is to provide the knowledge of how to correctly pronounce a word. When the visual input lexicon module and/or the semantic system is activated, correct pronunciation of a word is retrieved (*ibu* – `I bh U` or `i: bh U`).
- **Phoneme Level** – This module can be regarded as a short-term storage to hold the phonemes in the interval between speech output lexicon and speech production. In other words, this module holds the phonemes while the conversion of phonemes to articulatory movements took place. This module is especially needed when one tries to read aloud long, complicated words such as "tyrannosaurus" in English or *menyelenggarakan* in BM.

These are all the normal readers' models of reading. Unfortunately for dyslexics, some of the links between the modules are impaired causing them tremendous difficulties in learning to read. Therefore, for the purpose of this study, the model proposed by Ellis (1993) is adapted taken into consideration some changes based on the reading and spelling of dyslexic children that are introduced into the model. With the reading model selected, the following section and subsections review related literature for consideration prior to modeling an ASR model.

2.5 ASR 1, 2, 3

Prior to modeling an ASR model for dyslexic children, a brief review on background of ASR is presented, which covers ASR in BM, types of ASR, ASR methods, and ASR models. To map ASR to reading, it is worth to look at the three types of ASR: *discrete* ASR that deals only with completely *isolated* individual words; *connected* ASR that requires a user to pause between *connected* words; *continuous* ASR that recognizes continuous speech where *word boundary* detection is a must (Zulikha & Abdullah, 2007). In the early days, ASR technology concerns only in recognizing isolated speech – discrete ASR. However, current demand of ASR technology in various fields requires it to move to recognizing continuous speech. For the purpose of this study, a revisit to discrete ASR is needed to model isolated word recognition as mentioned for word level reading.

2.5.1 ASR for BM

Before going deeper into ASR types and methods, it is worth to mention that ASR has been around for BM speech recognition. The ASR is mainly focusing on the techniques for isolated digit recognition. The digits are "zero" to "nine" spoken in BM as in *kosong*, *satu*, *dua*, ..., *sembilan* (Md Sah, Dzulkifli, & Sheikh Hussein, 2001; Syed Abdul Rahman, Salina, Aini, Khairul, & Mirvaziri, 2007). ASR has also being used for speaker recognition by Sheikh Hussein, Ahmad, Zulkarnain, Rahman, and Lim (2000) for example. However, research in this field is limited and not as much as in English and other European languages. Even so, it is positively progressing from various works in digit recognition to BM syllable recognition to telephony-related vocabulary in BM (Ting, Jasmy, Sheikh Hussain, & Cheah 2001; Ting, Jasmy, & Sheikh Hussein, 2001; Zulikha & Abdullah, 2007). However, the above mentioned works concerned with adult speech, male and female, which differ significantly from that of children. Noteworthy, most of the works cited used HMM for recognition and are of discrete type of recognition as discussed in the following section.

2.5.2 Discrete vs. Continuous ASR System for Dyslexic

Discrete ASR involves recognition of words spoken one at a time whereas continuous ASR deals with continuous speech that obviously contains more than one word. Therefore, word boundary detection is not necessary in discrete ASR. On the contrary, continuous ASR is more complicated as it needs to carefully detect word boundaries for continuous speech. The discrete and continuous ASR differences are as listed in Table 2.2.

Table 2.2. The different effects of discrete ASR and continuous ASR adapted from Higgins and Raskind (2000).

Discrete ASR	**Continuous ASR**
No word boundary detection	Word boundary detection is necessary
Words appear one-by-one on screen	Words appear all at once in a sentence
Miss-segmentation is easier to detect and correct	Miss-segmentation is harder
Provides multi-sensory experience – simultaneous presentation of words	Does not provide multi-sensory experience – words take longer to appear on screen

Although ASR technology has the remedial potential for teaching and correcting reading among dyslexic children, the impact towards reading development vary in terms of the skills affected when discrete and continuous ASR is being applied. One important feature of discrete ASR for dyslexic children is that it provides multi-sensory experience, an approach that is essential for enhancing the children's ability to read (Higgins & Raskind, 2000). Continuous ASR however does not provide this to the children. Previous studies (Higgins & Raskind, 2000; Raskind & Higgins, 1999)[3] have found that the effects of discrete ASR and continuous ASR are different as depicted in Table 2.3. Based on the argument put forward previously, multi-sensory experience is indeed important for dyslexic children to learn to read. To enable multi-sensory

[3] Note that the ASR applications used by Raskind and Higgins in their longitudinal study is of dictation purpose and not specifically for reading. Their intention is to see whether or not the ASR applications used for dictation could help improve reading. As expected, based on the results, the use of ASR applications for dictation gave a promising potential to improve reading as well.

experience discrete ASR is more suitable as it can improve the children's word recognition skills, phoneme-grapheme awareness, and spelling.

Table 2.3. Discrete vs. continuous ASR effects on various skills adapted from Higgins and Raskind (2000).

Skills	Discrete ASR	Continuous ASR
Word recognition	Improved	Improved
Spelling	Improved	No
Written output	Slower	Faster
Phonological deletion	Improved	No
Phoneme-grapheme	Improved	No
Memory span	Not so much	Improved

Discrete ASR is more suitable for spelling and for younger children because they can see the orthographic representation of the spoken word as they pronounce it (Higgins & Raskind, 2000). Discrete ASR also gives better measurements towards word recognition, spelling, reading comprehension, and phonological processing as compared to continuous ASR. The continuous version conversely, increased written output generation and memory span of those children. However, this does not closely related to reading ability. Nevertheless, it suggests that continuous ASR can compensate writing better due to faster production of texts since it does not process one word at a time. Unfortunately, it does not mean that continuous ASR could remediate writing in the form of pen and paper (i.e. having the children to write on their own). It can be concluded, based on these studies, that discrete ASR has a positive effect on skills needed to read properly such as word recognition, phoneme-grapheme, and spelling. In addition, discrete ASR provides an excellent opportunity for children to practice correct pronunciations of words since it focused on accuracy of the uttered individual word (Russell et al., 1996).

On the contrary, Russell et al. (1996) also suggested using continuous ASR for teaching reading because it accentuates fluency and comprehension, although this introduces problems such as robustness and child-machine interface. Like Russell and colleagues, Fairweather et al. (n.d.) and Mostow et al. (1994) agreed that continuous ASR is more suitable when teaching children to read. They assert that reading is easier when expressed in continuous mode rather than word-by-word uttering, which is also not a 'natural' way of reading. Even children when asked to read tend to say a few words at a time. Their works, Watch-Me!-Read and Project LISTEN respectively, focus on automated reading instructions or reading tutor that uses ASR as the key technology in realizing the applications[4].

Enabling comprehension through fluent reading is beyond the scope of this study. Comprehension can be achieved if and only if a child could read properly and fluently. For dyslexic children, the first step to proper and fluent reading is to teach them to learn to compensate their difficulties by improving word recognition skill and thus increase automaticity, which can be done through the exposure to the multi-sensory experience. This is why discrete ASR is more suitable than the continuous version.

---

[4] The applications have been around for some time and have helped many children in learning to read. Automated reading tutors act as a tutor by 'listening' to the children reading out loud and try to track and identify reading mistakes in order to give feedback. The 'listening' process is done through successful recognition of the children's speech. The most important element of an automated reading tutor is to be able to give suitable feedback to users. Feedback is needed to respond to any reading mistakes or mispronunciations as well as to provide tracking, encouragement and suitable interventions to the children while they are reading.

2.5.3 ASR Methods

ASR technology has been around for decades and since then this technology has gained vast amount of interest and has been developed to serve many purposes such as computer controls, security, telecommunications, and education to name a few. With the creation of a statistical method called the Hidden Markov Model (HMM), ASR technology has been growing more rapidly. Other methods used in ASR technology include template matching, acoustic-phonetic approach, and artificial neural networks (ANN).

2.5.3.1Template Matching

Template matching is a method suitable only for small vocabulary ASR system as it needs a large space for storing the templates (Markowitz, 1996). The templates contain speech data as sets of features. Each template corresponds to one word or phrase. The recognition is made by comparing spoken input, put into a template, with the stored template.

The main problem of this technique is the space it requires to store all templates for recognition process. In addition, template matching is unable to produce accurate recognition results. Another drawback is that it is unsuitable for handling dyslexic children's speech recognition since it cannot handle words that share similar sounds (phoneme) which is always a problem for a dyslexic because they are normally confused by similar-sounding words. Although template matching was a dominant ASR technique back in the 50's and 60's, it gradually changes to become somewhat unpopular choice of method when acoustic-phonetic recognition technique is introduced. Rabiner and Juang (1993) provide a thorough description on this method.

2.5.3.2Acoustic-Phonetic Approach

Unlike template matching that operates at word level, acoustic-phonetic recognition technique operates at phoneme level (Markowitz, 1996). This is an interesting property of acoustic-phonetic recognition technique where it only requires storage of all the phonemes in a language. For an ASR system that recognizes BM for example, a total of 33 phonemes are needed regardless of the vocabulary size. This efficient property has made acoustic-phonetic technique a popular choice for an ASR application and suitable for large vocabulary systems. Three important steps of this technique are feature extraction, segmentation and labeling, and word-level recognition as described by Markowitz (1996).

Feature extraction involves extraction of spectral patterns of speech input that is needed to identify distinct phonemes. The extracted features are then labeled and segmented where phoneme boundaries are determined to distinguish from one another creating a set of phoneme hypotheses. These hypotheses are being used to match with the words stored in the application vocabulary in word level recognition. This technique seems an interesting approach for an ASR application that handles phonological awareness training for dyslexic children where it requires the children to pronounce words with similar-sounding properties aiming at enhancing their phoneme-grapheme association ability. However, due to the properties and estimation capabilities a stochastic method has, as discussed in the following sub-section, acoustic-phonetic is no longer a popular method for speech recognition.

2.5.3.3Hidden Markov Model

Another excellent alternative is a statistical approach to recognition where it makes non-deterministic selections among alternatives. Non-deterministic selection means that the choices made are solely based on input characteristics and not by pre-specified input. Unlike template matching that matches spoken input directly with stored templates, this technique matches spoken input with statistical and probabilistic analysis in a network-like structure called the hidden Markov model (HMM). Figure 2.4 illustrates the structure of HMM.

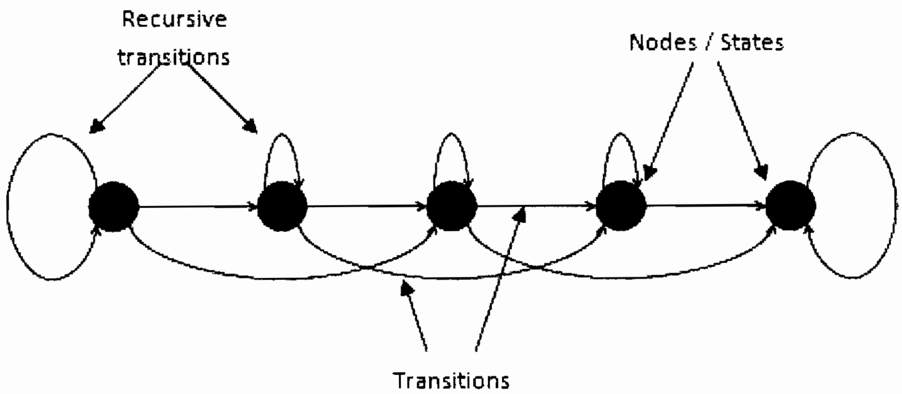


Figure 2.4. HMM structure consists of five nodes with their transitions and recursive transitions adopted from Rabiner and Juang, 1993.

HMM allows recognition both at word level and phoneme level. The states hold the statistics for a segment of a word whereas the transitions contain data that are used to make choice among the alternative states. Note that HMM allows recursive transitions i.e. a transition of a state back to itself. This property allows a prolonged phoneme/syllable uttered to remain in the same state through recursive transition. As an example, if a user prolonged ‘ba’ in *baca*, at least one recursive transition occurs.

Looking in dyslexic's perspective, while reading out loud, prolonged utterances of a segment of a word is common because they are either unsure of the pronunciation of a word or they cannot recognize the entire word. Moreover, dyslexic children have slower articulation and weak articulatory planning that caused somewhat prolonged utterances when trying to read (Fawcett & Nicolson, 2002). Nevertheless, HMM is a recognition technique that is accurate, flexible, and capable of automation. Putting these interesting properties of HMM into consideration HMM is indeed suitable for an ASR application for dyslexic children. Since 1990, HMM has become the dominant and most popular technique for ASR. HMM is discussed in some depth by Jelinek (2001) and Rabiner and Juang (1993).

2.5.3.4Artificial Neural Network Approach

Another popular choice of recognition technique is Artificial Neural Network (ANN). ANN is known for its excellent and powerful ability in classification. The basic architecture of an ANN comprises nodes assemble in an input layer, at least one hidden layer, and an output layer forming a network as shown in Figure 2.5. This network is akin of the neurological neural network of a human brain in terms of the structure and behaviour. Neurological neural network learns by sending information through connections called the synapses from neurons.

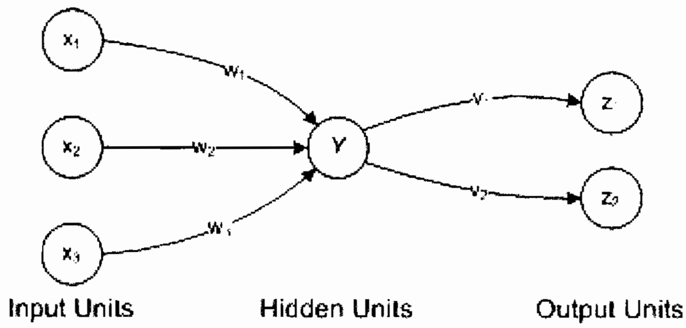


Figure 2.5. A multilayer neural network, illustrating an input layer, a hidden layer, and an output layer (source: Fausett, 1994).

Mimicking the neurological neural network operation, an ANN posses the ability to 'learn', which makes it powerful, by sending information from input layer to hidden layer to generate output in the output layer. The process of 'learning' is accomplished by *training* samples of data in a *supervised* or *unsupervised* mode. Supervised mode requires a target output to be specified prior to training whereas unsupervised network means that there is no target being specified. Examples of supervised and unsupervised ANN are Back Propagation and Kohonen Self-Organizing Map, respectively. Fausett (1994) provides an excellent reference to ANN and its underlying techniques.

2.5.3.5 Hybrid Method of HMM/ANN

Another interesting, excellent alternative of the above four techniques is a hybrid of HMM and ANN. The combined properties of HMM and ANN gives this technique a boost in recognition ability. The hybrid HMM/ANN approach is said to have outperformed the standard HMM in speech recognition making it a competitive alternative technique for ASR (Cosi, 2000; Franco, Cohen, Moran, Rumelhart, & Abrash, 1994; Renals et al., 1994; Rigoll & Willett, 1998; Trentin & Gori, 2001; Trentin & Gori, 2003; Yan, Fanty, & Cole, 1997). HMM/ANN hybrid recognition performance as opposed to the performance of pure HMM is demonstrated in Table 2.4. Even a simple ANN network can be used to estimate emission probability for HMM (Bourlard & Morgan, 1998).

Table 2.4. The WER percentage of recognition tasks using HMM and HMM/ANN performed in various domains.

Source	Word Error Rate (WER)		
	HMM	**HMM/ANN**	
Renals, et al. (1994)	11.0 %	5.8 %	ANN technique – Multilayer Perceptron (MLP)
	3.8 %	3.2 %	Interpolation of HMM and HMM/ANN (MLP)
Franco et al. (1994)	7.0 %	6.3 %	ANN technique – MLP
Yan, Fanty, & Cole (1997)	5.0 %	4.9 %	Feed-forward net, no specific technique
Trentin & Gori (2003)	9.97 %	9.80 %	Bourlard & Morgan's architecture
		6.10 %	Trained with Soft Weight Sharing Maximum-Likelihood (SWS-ML)
		6.75 %	Trained with Bayes Theorem
		5.35 %	Trained with Maximum a Posteriori (MAP)

These studies, although conducted years back, presented the results of recognition accuracy for both HMM and the hybrid methods for comparison measured in WER. It can be concluded that HMM/ANN method perform better than that of pure HMM. In addition, it is also acknowledged that the hybrid method is more efficient in terms of CPU and memory run-time processing (Bourlard & Morgan, 1998; Cosi, 2000; Trentin & Gori, 2001).

Since recognition accuracy is of concern, the ASR technique that best suites the purpose of this study should meet the criteria such that: it should be discrete ASR (single word reading and for multi-sensory effect); it should be able to handle children's speech, which very much differs from adults; and it should be able to cope with prolonged pronunciations of a word or certain parts of a word. As reviewed above, since HMM is a dominant technique in speech recognition, which major ASR tools are based on, coupled that with the power of ANN, the hybrid technique is thus chosen for this study. With the potential technique defined, it is worth to look at the potential ASR models to be adapted for improvement to further tuned the model for dyslexic children as discussed next.

2.5.4 ASR Models

ASR technology basically follows its basic components and process as illustrated in Figure 2.6. The speech is fed into a series of components before it can be recognized. The process includes preprocessing of the speech signal, feature extraction to extract significant features and speech recognition.

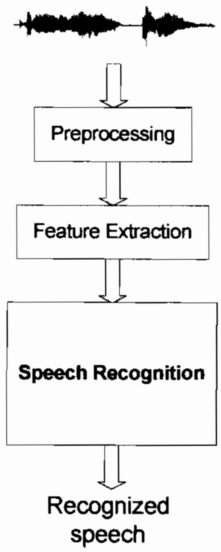


Figure 2.6. The general components involved in speech recognition.

The first fundamental process as depicted in the model is preprocessing of the speech signal. Preprocessing in general, involves filtering the speech signal and detecting the start and end point. Speech signal is later fed to feature extraction to extract relevant features for recognition purposes. The features can be referred to as a representation of a spectral envelope of the speech signal. The relevant features are then used as input for ASR to start the training process for recognition purposes. What makes a model different form another is the methods or techniques used to recognize speech. However, most ASR-based applications are running on HMM architecture or model. This is due to the fact that HMM is indeed the dominant technique in ASR.

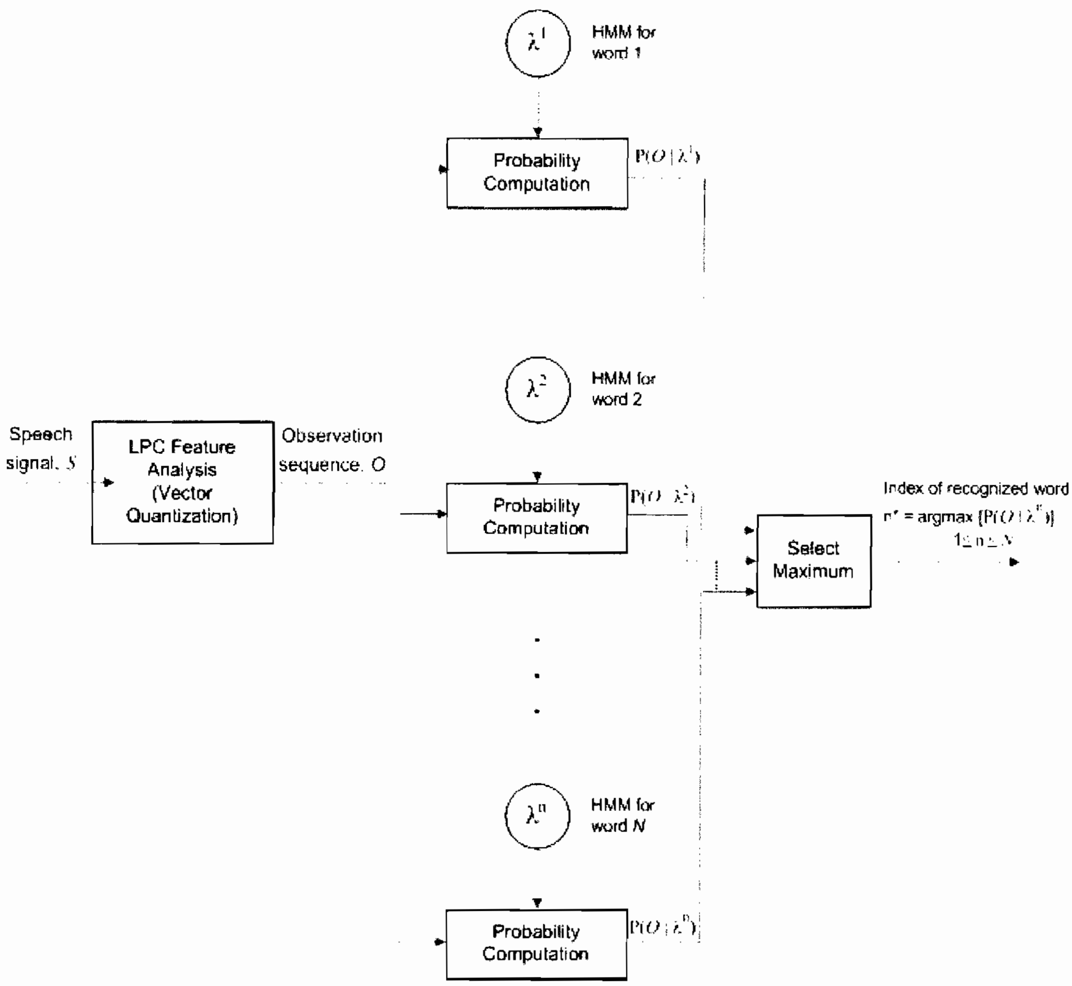


Figure 2.7. HMM architecture for isolated word recognition (source: Rabiner & Huang, 1993).

The HMM architecture depicted in Figure 2.7 illustrates the process of word recognition assuming a number of *N* words, using HMM. A HMM, denoted by λ^n, is built for each word by estimating the model parameters that can optimize the probability of the training set vectors for n^{th} word. Next, for any speech signal, *s* feature analysis is performed to obtain observation sequence, *O* corresponding to the speech. The observation sequence is then serve as input to compute probability of all possible HMM model built, λ^n given by $P(O \mid \lambda^n)$, $1 \leq n \leq N$. Normally, the probability estimation computation is performed either using Viterbi algorithm or Baum-Welch (forward-backward) algorithm (Trentin & Gori, 2001). Next, the probabilities computed is fed to the 'Select Maximum' process to obtain the highest probability estimation that determines which word the speech (input) is.

The difference between pure HMM and hybrid HMM/ANN architectures lies in the probability estimation methods used. Due to its undisputable classification ability as discussed in Section 2.4.3.4, ANN is used as an alternative for probability estimation in hybrid HMM/ANN (Hosom, Cole, & Cosi, 1999).

The hybrid architecture reviewed in the study is of frame-based recognition where at each frame, the ANN classifies phonetic-based categories based on the features in the context window (Shobaki, et al., 2000). Figure 2.8 depicts the hybrid HMM/ANN architecture.

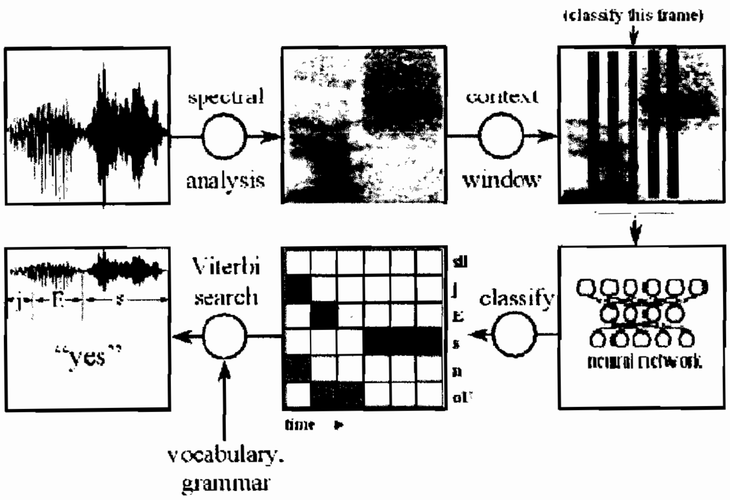


Figure 2.8. The illustration of the hybrid HMM/ANN architecture adopted from Shobaki et al. (2000).

In reference to the HMM/ANN architecture as illustrated in Figure 2.8, the recognition process starts from analyzing the speech by performing spectral analysis to obtain *frames*. Next is to compute features for each frame to be classified by the ANN and use its output together with pronunciation models and category information to determine the most likely word using Viterbi search (Hosom et al., 2006). Here, Viterbi algorithm is used to find the closest matching word based on the input rather than computing probability estimates, which is performed by ANN.

Adapting this architecture for the purpose of creating an ASR model for dyslexic children's reading is justified by the evidence of the recognition performance as compared to pure HMM method as presented in Section 2.3.4.5, as well as the advantages of using hybrid method rather than HMM alone, as discussed in the previous subsection.

The study adapts the hybrid HMM/ANN model and **emphasizes** on the "**vocabulary** and grammar" component and therefore is highlighted as one of the key element of ASR where the most frequent error patterns are represented. The next subsection justifies why the vocabulary or lexical model is regarded as an important component in an ASR model specifically for dyslexic children's reading.

2.5.5 Lexical Model and Context-dependent Modeling

With reference to Figure 2.8, the vocabulary represents the lexical model, which involves pronunciation models while the grammar constitutes to the language model. Although the components were not accentuated in Figure 2.8, they act as the key towards increasing the recognition accuracy (Huang et al. (2001) as cited in Noraini and Kamaruzaman, 2008). However, modeling the language model is beyond the scope of this study. Since the scope limits the study in discrete ASR for isolated word recognition, the language model serves to model only one word in a speech signal.

The focus is thus on the lexical model where the error patterns could be modeled. The lexical model is essential in ASR in which it acts as one of the *knowledge sources* for recognition (Hain, 2002; Rabiner & Juang, 1993). The representation and inclusion of the errors in the lexical model could drive to achieving gratifying recognition accuracy as suggested by previous studies working on ASR and reading for children (Banerjee et al., 2003a; Banerjee et al., 2003b; Fairweather et al., n.d.; Mostow et al., 1994; Mostow et al., 2002; Williams et al., 2000).

A lexical model can be defined as a "...combination of acoustic evidence so as to postulate words as specified by a lexicon that maps sounds into words (or equivalently decomposes words into sounds" (Rabiner & Juang, 1993). It involves the construction of a word from a series of phonemes that make up that particular word, i.e. a mapping from acoustic models and the words (Tsakalidis, Prasad, & Natarajan, 2009). For example, the phoneme sequence `bc bh tc th 0 l` (in Worldbet) is mapped to the word *betul*.

It is important to note that each pronunciation of a phoneme uttered even by the same person may slightly differ. In addition, each phoneme could be pronounced with relation to its previous or following phonemes. In BM for example, the phoneme `A` in *nyata* and *aku* is slightly different. The phoneme `A` in *nyata* is dependent on the digraph 'ny' where it sounds as though it is combined with the sounds of 'ny'. To accommodate such conditions, context-dependent pronunciation model is used as it could improve ASR performance significantly as asserted by Hosom (J.-P. Hosom, personal communication, March 31, 2009). Nevertheless, studies have demonstrated that context-dependent pronunciation model outperforms the independent version for various ASR (Husniza & Zulikha, 2009a; Husniza & Zulikha, 2010; Liu, Xu, Huang, Deng, & Li, 2000; Tsakalidis et al., 2009).

Theoretically, by utilizing the context-dependent model it is important to measure the performance of such model in relation with the dyslexic children's reading errors. Thus, the next section reviews evaluation methods and evaluation metrics pertinent to measuring the ASR model's performance.

2.6 Evaluation: The Recognition Accuracy

For the purpose of evaluation, an ASR recognizer is developed focusing on the lexical model as proposed. Due to the undisputable performance by the hybrid HMM/ANN method as reviewed, CSLU Toolkit (Sutton et al., 1998) is used to develop the ASR recognizer because it is the only available toolkit (for free for education and research purposes) that uses the hybrid method. Other available toolkits such as HTK and HMM Toolkit STK are based on HMM method only and therefore are not used in this study.

When measuring the performance of an ASR application, the most basic but essential question is how accurate is it? The answer is that the accuracy of an ASR application depends on the function for which the ASR is used, the purpose and criteria used to evaluate it, and the space in which the ASR evaluation is conducted (Mostow, 2006). According to Mostow (2006), words to be read make up a *text space*. In this space, the ASR is evaluated based on how accurate it can recognize each word as correct, misread or omitted from a sentence (omission is exclusively for continuous ASR application). Another space to evaluate an ASR is the *speech space* that consists of a sequence of words spoken by user that is heard by the ASR or human transcribers. In case of speech space, the ASR is evaluated in terms of how accurate it can transcribe words being read. Finally the *time domain space*, which aligns spoken words to time where evaluation can be done by comparing time-aligned output of ASR to that of the time-aligned transcript of what the user has read. Note that these spaces mentioned are for continuous speech recognition that deals with sentence-level reading. This study only focuses on the accuracy of whether the read word is correct or wrong and therefore shall evaluate the ASR model in the text space at word-level reading.

To measure the accuracy, in which whether the recognizer has accurately recognized spoken words, requires standard metric to be applied. Word error rate (WER) is a highly valid metric that is widely accepted and easy to use as claimed by Hagen et al. (2004). Examples of studies that use WER to measure and evaluate their ASR performances are by Russell et al. (1996) who managed to obtained equal error rate of 30%, Aist et al. (1998) with a low WER of 6.3% and 25.6% for KIDS corpus and reading tutor respectively, and Hagen et al. (2003) with average WER of 27.87%. Smith et al. (2002) in their quest to explore the available technology for prototyping and development of children's speech recognition product have found promising WER scores of 19.1% (female's speech) and 22% (male's speech).

Since this study concerns with dyslexic children's reading with phonetically similar errors, the optimum rate for WER is defined to follow that of state-of-the-art phoneme recognition rate between 70% to 75% as suggested by Hosom (J.-P. Hosom, personal communication, March 27, 2007). This means that the optimum rate for WER is defined to range from 25% to 30%.

Evaluating a recognizer in which the purpose is to recognize reading using only WER is not enough as it only provides the general recognition accuracy of how well the recognizer recognizes read speech independent from the target text or word (Mostow, 2006). Since the proposed recognizer is for recognizing the readings of dyslexic children, it would be very useful if the recognizer can perform well in detecting reading miscues as well. Mostow (2006) noted that "...WER gives zero credit for detecting reading mistakes unless it correctly recognizes the exact miscue the reader uttered...." Lee, Hagen, Romanyshyn, Martin, and Pellom (2004) support this notion as WER alone do not provide any diagnostic information. For this reason, miscue detection rate (MDR) is used as the metric to evaluate its performance in miscue detection. However, this does not mean that achieving lower WER is not important as various studies use this metric to evaluate their reading-related ASR performances (Aist et al., 1998; Hagen et al., 2004; Li, Ju, Deng, & Acero, 2007; Russell et al., 1996; Smith et al., 2002).

The ability to recognize miscues is seen important and meaningful especially when measuring the performance of an ASR that is built for reading recognition purposes where detecting reading errors are essential. MDR is often the metric used to measure the ASR performance when reading recognition is of concern (Lee et al., 2004; Mostow, 2006; Mostow, Beck, Winter, Wang, & Tobin, 2002; Tam, Mostow, Beck, & Banerjee, 2003;). Therefore, MDR provides richer information as to how accurate an ASR is when recognizing read words especially for the reading-oriented ASR. As illustrated in Table 2.5, previous works reported their highest obtained MDR for reading-related ASR that range from 42.53% to 88.80% with most works concentrating on English language (Banerjee et al., 2003a; Duchateau, Wigham, Demuynck, & Van hamme, 2007; Hagen, 2006; Lee et al., 2004; Li et al., 2007; Liu et al., 2008; Tam et al., 2003).

Table 2.5. MDR and FAR of various ASR.

	User type	Language	Evaluation Metrics MDR (%)	FAR (%)
Banerjee et al. (2003a)	Normal children	English	42.53	2.90
Tam et al. (2003)	Normal children	English	58.64	2.92
Hagen (2006)	Normal children	English	73.00	3.00
Li, Deng, Ju, & Acero (2008)	Normal children	English	76.90	15.80
Li et al. (2007)	Normal children	English	76.93	15.15
Lee et al. (2004)	Normal children	English	80.00	36.40
Duchateau et al. (2007)	Normal children	Dutch	83.10	8.40
Liu et al. (2008)	Normal children	Mandarin	88.80	11.50

Referring to Table 2.5, the optimum recognizer for normal children is given by Hagen (2006) with 73% MDR and a low FAR of only 3%. Even though Li et al. (2007) achieved a 3.93% increased in MDR, their FAR is 12% higher. However, these rates are obtained for the recognition of normal children's reading.

Unlike normal children, dyslexic children produce phonetically similar errors to target words due to their phonological deficits. As an example, they might misread *aku* as *aki* or *pin* as *pen* or *bapa* as *pada*. Note that these errors occur due to their difficulties to differentiate phonetically similar sounds. For ASR, recognizing phonetically similar vocabulary is a challenge because its limited ability in recognizing fine phonetic details, for example recognizing the presence and absence of a letter in a word (J.–P. Hosom, personal communication, May 14, 2008). Due to the phonetically similar vocabulary, the recognizer could mistakenly flagged correct readings as incorrect. This situation is regarded as *false alarm*. Thus, to measure the performance of a reading-oriented ASR, MDR is often measured together with false alarm rate (FAR) as illustrated in Table 2.5. Thus, FAR is the "percentage of correctly read words **rejected** by ASR" (Mostow, 2006). Therefore, MDR and FAR provide richer information as to how accurate an ASR is when recognizing read words especially for the reading-oriented ASR.

It is important to acknowledge that the performance of a speech recognizer depends on many aspects such as the amount of training data, vocabulary size, and type of words to name a few. Comparing the performances is tricky as each recognizer presented in Table 2.5 uses different training data, amount of data for training, different test data and even different languages, language models, and lexical models (Lee et al., 2004; Duchateau et al., 2007). Hence, the results of various recognizers presented are to illustrate the general idea of current available recognizers' acceptable MDR and FAR.

2.7 Chapter Summary

For the purpose of proposing an ASR model for dyslexic children's reading of single, isolated words in BM, it is worth looking into the theory that lies behind dyslexia and reading difficulties, that is the phonological-core theory. Originated from this particular theory, which is supported by many evidence from previous studies as discussed, the phonological impairment causes various reading and spelling errors. These errors are adapted in this study due to the cause-effect relationship presented – what problem causes what errors. Specifically, phonological deficit causes such errors to appear in dyslexic children's reading and spelling. Spelling is also taken into consideration because reading isolated words does involve spelling skill. The errors are adapted in the context of BM that differs from English. The errors are used to fine-tune a reading model for dyslexic children.

The reading model, which highlights the broad agreements between existing reading models, is chosen for the adaptation of the errors that is focusing on cognitive processes involved in word recognition. The errors and the reading model is connected to an ASR model for they provide vital information such as the dominant reading errors that are incorporated into the ASR model as vocabulary. The vocabulary is incorporated into the ASR lexical model, which is an important element to any ASR.

The ASR model considered for adaptation is of HMM/ANN hybrid method since it has been proven that this method is capable to outperform HMM method, which dominates ASR technology thus far. With that, this chapter has reviewed the related literatures towards achieving objective a, b, c, d, e, and f of Chapter 1, Section 1.4. The next three chapters present more comprehensive discussions to achieve the specified objectives and the final chapter concludes the thesis.

CHAPTER 3

METHODOLOGY

3.0 Introduction

The previous chapter reviews literatures related to the research objectives. An extensive body of literature is reviewed in order to build knowledge and understanding of the background domain of study. The significant review led to suitable methods selected to perform research activities to achieve each of the objectives specified in Chapter 1 Section 1.4. Hence, this chapter continues using the chosen methods. The methodology outlines four tasks to be performed as the following: 1) Task 1 – Speech collection of dyslexic children's reading; 2) Task 2 – Dyslexic children's reading model and ASR model creation; 3) Task 3 – ASR recognizer development; and 4) Task 4 – ASR recognizer's evaluation.

Task 1 involves collecting primary data of dyslexic children's reading aloud single, isolated words in BM. Therefore, the choice of words to be selected as reading material or stimuli is of significant value. In line with the first objective, which is to collect vocabulary in BM, methods such as unstructured interview and observation are selected. Interviews with special education teachers and private dyslexia tutors were conducted in an unstructured, informal manner. The informal interview led to initial findings of which BM words are suitable and why. Observation, on the other hand, is performed to obtain the reading pattern of the dyslexic children participated in the study. The observation is conducted in a total of three months until significant patterns emerge from the activity.

Task 2 is about the creation of models, both reading and spelling model for dyslexic children and the ASR model that is specifically designed for these children. In this phase, data analysis is performed to extract significant reading and spelling error types and patterns from the data collected. The patterns serve as essential input for both the reading and spelling model of dyslexic children as well as the ASR model.

Once the data are analysed and the models are developed and proposed, next two tasks involve development and testing and finally evaluation of such model. The method for model evaluation is by prototyping. Rapid application development (RAD) is employed for building the ASR recognizer of the proposed ASR model. Evaluation is performed to evaluate the recognizer in terms of its recognition accuracy using WER and MDR as metrics as discussed in Chapter 2, Section 2.6. The four tasks are depicted in Figure 3.1.

Accordingly, this chapter is organized based on the tasks as mentioned above. Section 3.1 describes methods for speech data collection from selecting vocabulary and suitable dyslexic children to conducting data collection. The following section, Section 3.2 delineates methods for data analysis which findings lead to the creation of reading and spelling model and the ASR model to be proposed. Section 3.3 outlines the development of an ASR recognizer and Section 3.4 describes the evaluation. Finally, Section 3.5 concludes the chapter.

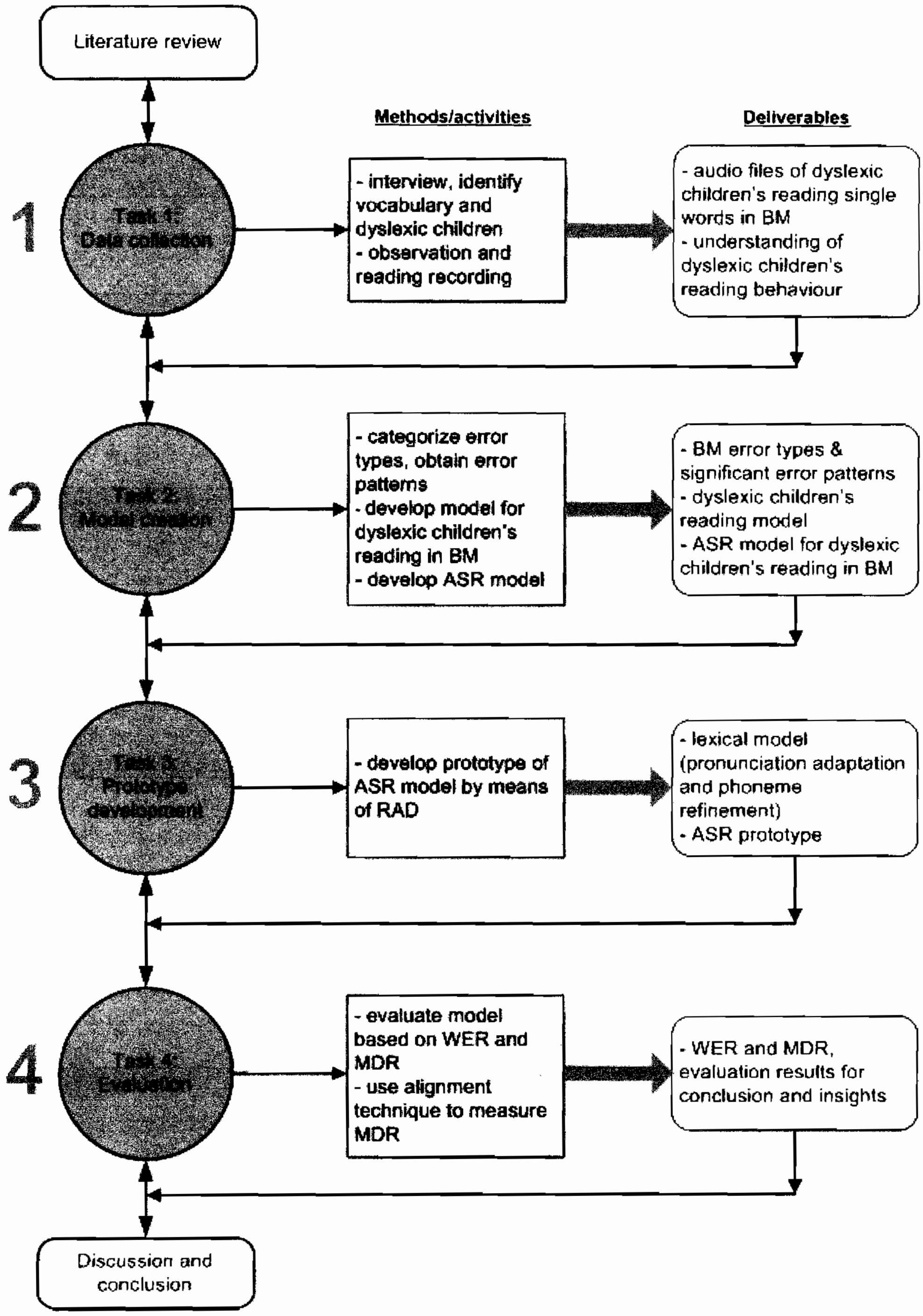


Figure 3.1. The methods and deliverables encapsulated in the research framework.

3.1 Task 1: Data Collection

Children who have dyslexia have somewhat different and rather unique reading patterns as compared to their normal peers. Their reading patterns resemble the reading patterns of children who are *at least* two years below their age. Sadly, some would read very poorly – years below their chronological age. A boy, given a persona name Aiman, in this study, to take as an example, is a bright, courteous, and friendly 14 year-old but his reading resembles that of a 9 year-old child with a lot of reading mistakes. He keeps on adding extra syllables to any word and thus hinders his reading ability not to mention his reading comprehension. To find out more on dyslexic children's reading patterns, two techniques are used mainly, an extensive review of literature and an informal interview with teachers and facilitators.

An extensive review of literature is the key in identifying spelling and reading error patterns as well as the underlying theories that support such patterns. As significant as the reading errors, spelling errors are the essential 'ingredients' for analyzing single, isolated word reading. This is due to the interdependencies between the two skills (i.e. reading and spelling) as discussed.

To know dyslexic children and their reading nature better, discussions through an informal interview is performed. The discussions involved special education teachers from public schools and facilitators from private organizations that offer special classes for dyslexic children. The two participated schools are Sekolah Kebangsaan Taman Tun Dr Ismail (2), Kuala Lumpur and Sekolah Kebangsaan Jalan Datuk Kumbar, Alor Setar. The organizations are Bureau of Learning Difficulties (BOLD), Penang and Seberang Jaya as well as Persatuan Disleksia Malaysia, Kuala Lumpur and USJ 6, Subang Jaya.

3.1.1 The vocabulary

The population of concerned in this case is the words that commonly appear in Year 1, 2, and 3 BM text books and *Buku Panduan Pelaksanaan Program Pemulihan Khas: Masalah Penguasaan 3M* (Jabatan Pendidikan Khas, 1999). Note that these words are representations of 23 syllable pattern groups corresponding to BM spelling patterns as specified in the syllabus. The syllable patterns also represent level of decoding complexity for dyslexic children to decode.

A *random cluster* method is used to select the vocabulary since it is faster and thus provides efficient means to select vocabulary from thousands of words in the text books. Words that match each group of the syllable patterns in BM are randomly selected from the text books and the *Buku Panduan Pelaksanaan Program Pemulihan Khas: Masalah Penguasaan 3M* (Jabatan Pendidikan Khas, 1999). The selected words from the books are included in their corresponding clusters. A cluster can be regarded as a small scale representation of a population, in this case, the words that represent each syllable pattern. Words within the syllabus that commonly appear in the text books are tabled according to their correspondent clusters. Here, the clusters are the 23 syllable patterns in BM taking into consideration the prefixes and postfixes as listed in the *Buku Panduan Pelaksanaan Program Pemulihan Khas: Masalah Penguasaan 3M*. Table 3.1 illustrates some of the common words listed in their corresponding clusters (denoted by the consonant, C and vowel, V).

Table 3.1. Example of clusters of common words within level 1 syllabus, tabled accordingly in random cluster sampling technique.

V+CV	CV+CV	CV+CV with digraph	CV+CV with diphthong	CVC
aku	*kaki*	*bunga*	*gurau*	*dan*
itu	*saya*	*tanya*	*kilau*	*wah*
ini	*baca*	*sunyi*	*wahai*	*bin*
ibu	*bapa*	*punya*	*kelui*	*jam*
apa	*suka*	*ngeri*	*ceria*	*pen*
...	...	...	...	...

V+CVC	V+CVCC	CV+CVC / CVC+CV	CV+CVCC	CVC+CVC
ayah	*abang*	*datuk*	*sayang*	*kerbau*
adik	*udang*	*nenek*	*sering*	*keldai*
ayat	*inang*	*tahun*	*burung*	*gambar*
umur	*ulang*	*rumah*	*kucing*	*hampir*
ikut	*ulung*	*pergi*	*bawang*	*rambut*
...	...	...	...	...

CV+CVC/ CVC+CV with diphthong & digraph	CVC+CVC with digraph	Paired vowels (*vokal berganding*)	CV+CV+ CV	CV+CV+ CCV
pantai	*kangkung*	*dia*	*selesa*	*telinga*
sangat	*bengkung*	*zoo*	*kereta*	*pelangi*
langkau	*pinggang*	*air*	*dahulu*	*kenanga*
pandai	*lenggang*	*tiup*	*kepala*	*belanga*
jangan	*tunggang*	*suap*	*negara*	*berenga*
...	...	...	...	...

CV+CV+ CVV	CV+CV+ CCVV	CVC+CV+ CVC	CV+CV+ CVCC	CV+CVC+ CVCC
senarai	*perangai*	*kumpulan*	*binatang*	*penumpang*
merayau	*meringai*	*bercakap*	*terowong*	*memancing*
serunai	*jerangau*	*cendawan*	*belakang*	*melambung*
belalai	*melangai*	*jemputan*	*belalang*	*menendang*
kemarau		*maklumat*	*meradang*	*berunding*
...		...	...	...

CVC+CV+ CVCC	CVCC+ CV+CVC	CV+CVCC +CVC
pendatang	*bungkusan*	*meningkat*
tempurung	*jangkitan*	*menangkap*
pemborong	*bingkisan*	*melanggar*
pembayang	*panggilan*	*perangkap*
tembelang	*bangsawan*	*selongkar*
...	...	...

Within each cluster, an exhaustive list is produced comprising the randomly selected words as mentioned. For illustration, Table 3.1 depicts some of the words to represent each cluster. Next, for each cluster, five words are therefore chosen randomly according to the *two-stage random sampling*. The syllable pattern CV+CV+CCVV, which contains words that are not so commonly appear in the text book, has only four samples as listed in the Buku Panduan Pelaksanaan Program Pemulihan Khas (Masalah Penguasaan 3M). Thus, the words *perangai*, *meringai*, *jerangau*, and *melangai* are all selected to represent the CV+CV+CCVV cluster. A total of 114 words are chosen to represent all clusters. Table 3.2 illustrates one example of word selected to represent each syllable pattern.

Table 3.2. Examples of words for each of the syllable patterns of consonant- vowel categories and used as stimuli for data collection.

	Syllable Pattern	Word		Syllable Pattern	Word
1.	V+CV	*ibu*	13.	Paired vowel (*vokal berganding*)	*puas*
2.	CV+CV	*suka*	14.	CV+CV+CV	*kepala*
3.	CV+CV with digraph	*nyata*	15.	CV+CV+CCV	*telinga*
4.	CV+CV with diphthong	*wahai*	16.	CV+CV+CVV	*senarai*
5.	CVC	*dan*	17.	CV+CV+CCVV	*perangai*
6.	V+CVC	*umur*	18.	CVC+CV+CVC	*kumpulan*
7.	V+CVCC	*abang*	19.	CV+CV+CVCC	*binatang*
8.	CV+CVC / CVC+CV	*rumah*	20.	CV+CVC+CVCC	*penumpang*
9.	CV+CVCC	*sayang*	21.	CVC+CV+CVCC	*pendatang*
10.	CV+CVC / CVC+CV with digraph & diphthong	*sangat*	22.	CVCC+CV+CVC	*panggilan*
11.	CVC+CVC	*keldai*	23.	CV+CVCC+CVC	*melanggar*
12.	CVC+CVC with digraph	*kangkung*			

A diphthong refers to "...a gliding monosyllabic speech sound that starts at or near the articulatory position for one vowel and moves to or toward the position for another" (Rabiner & Juang, 1993). Examples include the sounds of 'ai' in *wahai* and 'oi' in *amboi*. A digraph is "a group of two successive letters whose phonetic value is a single sound (as 'ea' in "bread" or 'ng' in "sing") or whose value is not the sum of a value borne by each in other occurrences (as 'ch' in "chin" where the value is \t\ + \sh\)" as defined in the Merriam-Webster online dictionary (http://www.merriam-webster.com/dictionary). Examples in BM include words like *nyata* and *ngilu* where a single sound is represented by two successive letters 'ny' and 'ng'. The syllable patterns, although some are rather complicated, is sufficient to test the children's decoding ability for word recognition as suggested by their teachers. These words when presented as stimuli served as the key towards discovering the spelling and reading error pattern needed to develop the reading model as well as the ASR model. The overall process of selecting sample words to represent each cluster and build the vocabulary is depicted in Figure 3.2.

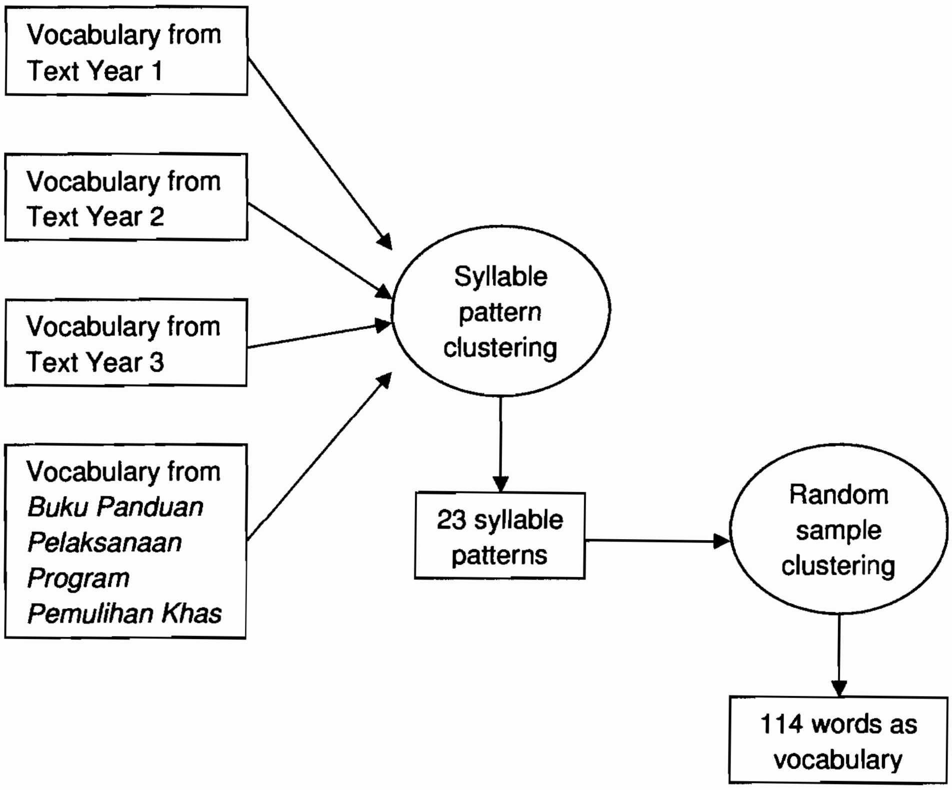


Figure 3.2. The process of selecting the words to serve as vocabulary and stimuli.

3.1.2 The participants – Dyslexic children

Ten dyslexic children with the same reading age, regardless of their chronological age are chosen for this study. In the two schools there are only a total of 12 dyslexic children who read at word recognition level, which is the focus of this study, based on their teachers' suggestions. Two children are ruled out because one did not know the alphabets, hence leading him towards a much lower word recognition level of reading, and the other is due to her lack exposure to BM. Hence, having ten dyslexic children is believed to be representative as to dyslexic schoolchildren and to be able to attribute to a generalized dyslexic children's reading model as well as ASR model as shown in the previous studies concerning ASR and reading (see Chapter 2). Besides, this reading model has a major influence in the ASR model's performance when tested later on. All ten dyslexic children who participated in this study have been suggested by their teachers who are acquainted with their reading level. Few criteria fall for consideration in selecting suitable dyslexic children for observation and recordings for this study. The criteria are used for screening purposes and are listed as follows:

- **A dyslexic child should be formally diagnosed** – this can only be conducted either by psychologists or doctors, or through diagnosis using an instrument provided by the government.
- **A dyslexic child must read at similar reading age regardless of chronological age** – a child cannot be included in the study should his/her reading age is above or lower than the chosen level.
- **A dyslexic child should have similar exposure to BM language** – children without exposure to BM are most likely unable to read in this language. This is to ensure that the children participated in the study have enough exposure of BM in their environment that they can communicate fluently and understand instructions given in BM, as well as being regularly exposed to reading material in BM. Unfortunately, despite all that, they still fail to read properly. Language exposure is regarded as an environmental factor of dyslexia for multiple language users. Figure 3.3 illustrates a causal model by Dyslexia Association Singapore, DAS (2003) which include language exposure as one of the environmental causal factor.

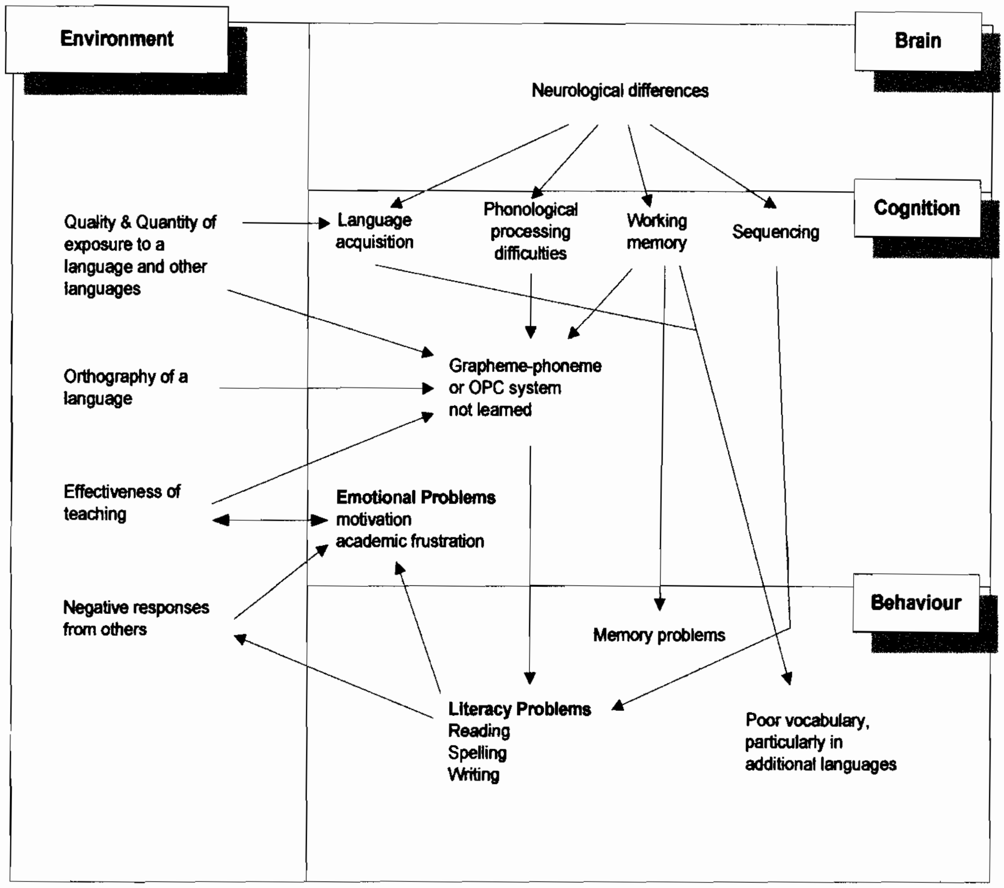


Figure 3.3. DAS's causal model of dyslexia, illustrating the various environmental factors in a bilingual or multilingual community, adopted from DAS (2003).

- **A dyslexic child must have knowledge of the alphabet** – a child must not be included in the study should he/she does not know the letter. This indicates a severe symptom of dyslexia that leads to a much lower reading age and so the child is not suitable for data collection purposes.
- **A dyslexic child must have reading problem at word level (isolated words)** – a child who can read quite fluently in sentences and paragraphs are not included since they can fluently recognize and pronounce isolated words. Their target would be comprehension rather than word recognition.

Worth to mention that data collection and observation involving young children, whose chronological age ranges from 7 to 14 years old, is rather challenging and time consuming especially for dyslexic children. Not all of them cooperate with 100% commitment when asked to read aloud during the recording sessions that later affect their reading performances, which vary every session. Some are playful and talkative and often refuse to follow instructions and some are rather quiet and thus read with very soft voice. All of these affect the time to record, the quality of the readings, and the quality of speech produced when reading aloud.

3.1.3 Primary data collection

In the previous two subsections, methods of selecting vocabulary and choosing suitable dyslexic children as participants are discussed. Next is to deliberately set out the methods involved in primary data collection for this study. An ethnographic technique specifically observation is performed in this fieldwork to obtain spelling and reading error patterns. Since dyslexic children have inconsistent reading each session, observation is conducted for seven sessions to retrieve their patterns and obtain the most frequent patterns. Note that for each participant, the sessions are conducted in different days which mean that only one session is conducted in one day. Two simultaneous activities have been performed namely observations and recordings of the children's reading.

Observation is performed in order to obtain the spelling and reading error patterns of the children when they are prompted with various selected BM words. Recording is performed to capture the reading in audio form. The audio files recorded serve two purposes: (1) to enable valid and careful analysis of the errors observed and enable re-listening to the recordings; (2) to use in the development of the ASR model for the purpose of training and testing the ASR recognizer and evaluation.

For recording purposes, *SpeechViewer* tool of CSLU Toolkit (Sutton et al., 1998) is used to record the children's reading at 16k samples/sec. A standard headphone with microphone is used in order to reduce background noise as much as possible. The following are the steps involved in the observation and reading recording session.

Step 1: Prompt a word from a list

Step 2: The child read the word – listen carefully, give constant encouragement and praise if the word is correctly pronounced.

Step 3: Record the reading simultaneously

(repeat Step 1 and 3 if the recorded reading is insufficient for later use)

Step 4: Take note if any, especially when the reading errors produced are of prominent pattern or out of the ordinary for that individual (as observed in previous sessions). For example, DC-10 always read by adding the postfix '-an' and this pattern of errors is consistently observed only for this participant. Other participants rarely made such mistakes.

Step 5: Repeat Step 1 to 4 until all 114 words are read

(some participants are only able to read half or less of the words in a session and so the recording continues the session after)

Noteworthy, recording is performed on individual basis in a reasonably quiet room somewhere in the schools, if the rooms are available. The length of each recording session differs depending upon each participant's ability and willingness to perform the procedure on that particular session. Their ability to concentrate and read prompted words also varies each session. They can cooperate and be really attentive towards recording in one session, and be the opposite in the next session. In addition, not all of them give a 100% commitment each session. When this happen, they usually refuse to follow instruction and thus read in an unwilling, awkward reading that often produces incorrect pronunciations of words. This situation hinders the recordings to progress smoothly and thus increases the time needed to complete all 114 words for seven sessions of recording for each participant. The delays lead to a time consuming process especially when a slower dyslexic reader is involved. Slow dyslexic readers fail to finish reading the stimuli in one recording session. Sometimes, they require up to three sessions to complete all 114 words. Thus, the entire data collection took three months to complete all ten participants. Table 3.3 illustrates one example of transcribed readings for the word *sunyi* recorded (Rec.) in seven reading recording sessions for all ten participants, DC1 to DC 10. For a complete set of transcribed readings for all 114 words, refer to Appendix B.

Table 3.3. The readings of *sunyi*, which have been transcribed into corresponding spellings according to how they were pronounced.

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
	DC1	*sunyi*	*sunyi*	*sunyi*	*sunyi*	*sunyi*	*sunyi*	*sunyi*
	DC2	*sunyi*	*sun wi*	*sun ya*	*sun yi*	*sunyia*	*sunyi*	*sun yi*
	DC3	*sembunyi*	*bunyi*	*sembunyi*	*sembunyi*	*sembunyi*	*sembunyi*	*sembunyi*
	DC4	*suni*	*suni*	*sunyi*	*sunyi*	*sunyi*	*sunyi*	*sunyi*
	DC5	*suni*	*suyi*	*suni*	*sunyi*	*suni*	*suni*	*suni*
sunyi	DC6	*sun yi*	*sun yi*	-	-	-	-	-
	DC7	*sunya*	*sunnya*	*suni*	*sunyi*	*sunyi*	*sunyi*	*sunyi*
	DC8	*suya*	*suna*	*sunyi*	*sunya*	*sunyi*	*sunyi*	*sunyi*
	DC9	*saya*	*sunyi*	-	-	-	-	-
	DC 10	*suni*	*suyi*	*sunyi*	-	-	-	-

Referring to Appendix B, looking at the recording patterns it can be concluded that DC6, DC9, and DC10 did not complete all seven sessions of reading recordings due to their ability to read, which is very slow and exhaustive and thus time consuming. Note that the recordings are performed in seven sessions to allow for their readings and spelling error patterns to immersed and therefore identified. Even though they did not complete all seven sessions, their reading error patterns are obvious when considering transcriptions of other target words. DC6 clearly have 'blending' problem, which is the difficulties to blend together correctly spelled (and sounds) syllables in a word. DC9 often reads with substitutions and simply guessing the words, and DC10 normally adds extra syllable at the end of the words (example such as *adangan* for *abang*).

3.2 Task 2: Model Creation

The data collected serve as the basis towards building a reading model for the children. The model is derived from the observation and audio recordings of isolated word reading that have been carefully analysed. Since both models require analysed data as essential input, the activities involved in task 2 are as follows:

- Data analysis
- Improvement of reading model
- Improvement of ASR model

Data analysis is performed to obtain reading and spelling error types and patterns in the BM context. The pattern emerged is used as important element in the reading model to be proposed as well as for improvement of the chosen ASR model. Discussions on existing reading models and ASR models are presented in Chapter 2 Section 2.3.

3.2.1 Data analysis

The analysis is conducted to discover and classify any errors immersed from the data into suitable categories adapted from Sawyer et al. (1999). The analysis constitutes to two procedures: 1) Re-listening of the recordings and simultaneously transcribing the data (read words) into corresponding spelling in BM; and 2) grouping the errors into designated categories.

The first is concerned with categorizing and assigning spelling and reading errors into specific phonological-based error types adapted from Sawyer et al. (1999) as discussed in Chapter 2. Although BM uses the same orthographic representation as in English, everything else about the language is entirely different from the phones of the letters to the spelling system and grammar that a BM sentence holds. Hence, the categories presented in Sawyer and her colleagues are adapted taking into consideration the possible BM-based errors that cannot fit into the available categories. Such errors include the reading errors that emerged from confusing the syllable division in a word. Table 3.4 presents an example of error classification for the word *sunyi.*

Since the categories listed by Sawyer et al. (1999) are based on English spelling scheme, two of the categories are omitted to suite the spelling of BM. The categories are "add final e" and "omit final e". In addition, five new categories are introduced to incorporate errors which are based on BM spelling system such as "syllable division confusion" which can obviously be seen in read words such as "bun ga" for *bunga* and "tan ya" for *tanya.* A list of these added categories are shown in Table 3.5 marked by '*' symbol.

Table 3.4. Classification of errors for read words, which have been transcribed into corresponding spellings.

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
sunyi	DC1	*sunyi*	*sunyi*	*sunyi*	*sunyi*	*sunyi*	*sunyi*	*sunyi*
	Error	-	-	-	-	-	-	-
	DC2	*sunyi*	*sun wi*	*sun ya*	*sun yi*	*sunyia*	*sunyi*	*sun yi*
	Error	-	SDC, SC	SDC, SV	SDC	AV	-	SDC
	DC3	*sembunyi*	*bunyi*	*sembunyi*	*sembunyi*	*sembunyi*	*sembunyi*	*sembunyi*
	Error	SW	SC	SW	SW	SW	SW	SW
	DC4	*suni*	*suni*	*sunyi*	*sunyi*	*sunyi*	*sunyi*	*sunyi*
	Error	OC	OC	-	-	-	-	-
	DC5	*suni*	*suyi*	*suni*	*sunyi*	*suni*	*suni*	*suni*
	Error	OC	N	OC	-	OC	OC	OC
	DC6	*sun yi*	*sun yi*	-	-	-	-	-
	Error	SDC	SDC					
	DC7	*sunya*	*sunnya*	*suni*	*sunyi*	*sunyi*	*sunyi*	*sunyi*
	Error	SV	N, SV	OC	-	-	-	-
	DC8	*suya*	*suna*	*sunyi*	*sunya*	*sunyi*	*sunyi*	*sunyi*
	Error	N, SV	OC, SV	-	SV	-	-	-
	DC9	*saya*	*sunyi*	-	-	-	-	-
	Error	SW	-					
	DC 10	*suni*	*suyi*	*sunyi*	-	-	-	-
	Error	OC	N	-				

*SDC – Syllable division confusion
SV – Substitute vowel
SW – Substitute with word
N – Nasals
SC – Substitute consonant
AV – Add vowel
OC – Omit consonant

Table 3.5. Categories of error types for analysis of read words error classification and the corresponding definitions.

	Categories of Error Types	Definitions
1.	Nasals (add/delete/substitute – *m* or *n*)	Addition, deletion, or substitute *m* or *n* (*makam* for *makan*; *maka* for *makan*) Note: substitution is of either between *m* and *n* or substitutes other vowel or consonant with *m* or *n*.
2.	Substitutes nasals for liquid	Substitution of a nasal for a liquid (*beni for beli*)
3.	Liquids (add/delete/substitute – *l* or *r*)	Addition, deletion, or substitute *l* or *r* (*beri* for *beli*) Note: substitution is of either between *l* and *r* or other vowel or consonant with *l* or *r*.
4.	Omitted consonants (excludes *m, n, l ,r*)	Errors of consonant omissions (*apa* for *bapa*)
5.	Substitutes consonants	Substitution of one consonant with another (*sarang* for *barang*)
6.	Adds consonants	Adding a consonant (*bapak* for *bapa*)
7.	Substitutes vowel	Substitution of one vowel with another (*umir* for *umur*)
8.	Omits vowel	Omission of a vowel (*merayu* for *merayau*)
9.	Incorrect sequence	Inappropriate sequence of letters (*paus* for *puas*)
10.	Substitutes word	Substitution of one word for another (*dengan* for *dan; tempat for pantai*)
11.	Reversals	Reversing the form of a letter (*b* for *d* or *p; m for w*)
12.	* Syllable Division Confusion	Confusing the division of syllables in a word (*bun|ga* for *bunga*)
13.	* Adds vowel	Adding a vowel (*sunyia* for *sunyi*)
14.	* Omits syllable	Omission of either front, middle or final syllable in a word (*ta* for *tanya*)
15.	* Adds syllable	Adding extra syllable to a word (*se<u>na</u>runai* for *serunai*) Note: only valid syllables are considered, invalid syllable e.g. *se<u>nne</u>runai* are not assigned to this type.
16.	* Substitutes vowel with consonant or consonant with vowel	Substituting a vowel with a consonant or vice versa (*g* for *a*) if: - substitution of vowel with consonant (excluding *m, n, l, r*) - substitution of consonant (including *m, n, l, r*) with a vowel

The categories serve as the fundamental component towards the development of the reading model. The analysis then continues to examine the most frequent error patterns. The error patterns are obtained by looking at the error categories made through observation and recording sessions for each of the participants.

Once the error type patterns are recognized, the most frequent error pattern is obtained by selecting the highest percentile. Noteworthy, the frequency and percentile calculated suggest which errors that most dominantly occurs in the participants' reading and spelling of the selected words. The results give "substitute vowel" as the most frequent error pattern as shown in Table 3.6 (adapted from Husniza & Zulikha, 2009b).

Table 3.6. Reading and spelling errors, in descending percentile, of dyslexic children reading aloud controlled vocabulary in BM (Source: Husniza and Zulikha, 2009b).

Error types	***n***	**%**
Substitutes vowel	1286	21.25
Omits consonants *	786	12.99
Nasals (*m, n*)	770	12.73
Substitutes consonants *	577	9.54
Omits vowel	511	8.44
Substitutes word	384	6.35
Adds consonants	363	6.00
Reversals	268	4.43
Incorrect sequence	224	3.70
Omits syllable	167	2.76
Liquids (*l, r*)	156	2.58
Substitutes vowel with consonant / consonant with vowel **	143	2.36
Substitutes nasals for liquid	124	2.05
Adds vowel	124	2.05
Syllable Division Confusion	94	1.55
Adds syllable	74	1.22

* excludes *m, n, l ,r*

** *if:- substitution of a vowel with a consonant (**excluding** m, n, l, r)*
*- substitution of a consonant (**including** m, n, l, r) with a vowel*

It is worth mentioning that the three most frequent errors obtained from this study replicate those of Sawyer et al. (1999) when children read single words in English. Unlike Sawyer et al. who present consonant omissions and substitutions collectively as the second largest category of errors, this study shows that consonant omissions and nasals are the second most frequent errors made when the children read aloud single words in BM with both reaching over than 12%. Consonant substitution, which is believed to have been related to "possible misperception of similar sounds" as claimed by Sawyer et al. does not involve substitutions of voiceless consonant. Substitutions of voiceless consonant letters appeared to be the biggest mistake which contribute to the number of substitution errors. Unlike English, BM has no voiceless consonant letters. However, there is similarity in terms of the results which exhibit that the most frequent errors made is of vowel substitutions. In addition, other similarity includes that consonant omissions and substitutions are both included in the most frequent errors. For both English and BM, vowels are represented by the letter 'a', 'e', 'i', 'o', and 'u'. Given the knowledge of what error pattern is dominant among the participant's spelling and reading, words that contain such patterns are therefore selected and the pronunciations are modeled using method as described in the following section.

3.2.2 Pronunciation model

The pronunciation model is implemented manually by producing hand coded transcriptions of the selected words' citations to their corresponding phones. Transcription can also be performed automatically but since these are all read speech, *manual transcription* is always better (J.–P. Hosom, personal communication, December 18, 2008) and thus sufficient as read speech pronunciations do not require considerations of the variability factor as in spontaneous speech such as accent. Note that because this involves a small vocabulary, which consists of only the selected words from the most frequent spelling and reading error patterns, the manual transcription approach is less complicated and therefore not costly. The words citations are transformed into their corresponding phones (sounds) using the Worldbet representation (Hieronymus, 1993).

Worldbet is an ASCII representation based on the International Phonetic Alphabet (IPA). Instead of just IPA symbols, Worldbet also include other phonetic symbols to represent other phones that exist in the world languages that are not present in English and European languages. Since this study is focusing on the spelling and reading errors in BM, Worldbet is a suitable representation and therefore chosen for this task. The following Table 3.7 illustrates examples of pronunciation models for *ibu, aku, dan, rumah,* and *kereta* in Worldbet represetation.

Table 3.7 Pronunciation models for word *ibu, aku, dan, rumah,* and *kereta* in Worldbet.

Word	**Worldbet**
1. ibu	`i: bc bh U or I bc bh U`
2. aku	`A kc kh U`
3. dan	`dc dh A n`
4. rumah	`9 U m A h`
5. kereta	`kc kh & 9 E tc th A`

The labeling method used in this study follows that of the CSLU Toolkit's Worldbet representation taking into consideration the diacritics symbols and allophones, which is in accordance with the second pronunciation labeling method using Worldbet as described by Hieronymus (1993). The first method is to follow the basic form of an ASCII representation of the IPA and the third labeling method take into consideration the labeling of the actual regions of allophonic production by the vocal apparatus.

3.2.3 Improvement of the reading model

The reading model of Ellis's (1993) are adapted (see Chapter 2, Section 2.4 for a review of reading models) to introduce a reading model of dyslexic children with additional attributes in the context of BM embedded in the original model. The additional attributes are contributed by the most frequent error patterns as shown in Table 3.4. The patterns included in the reading model are given by "substitute vowel", "omit consonant", "nasals" and "substitute consonant". These errors are similar to that of Sawyer et al. (1999) which gives vowel substitutions and combination of consonant-based errors as the most frequent errors made by dyslexic children.

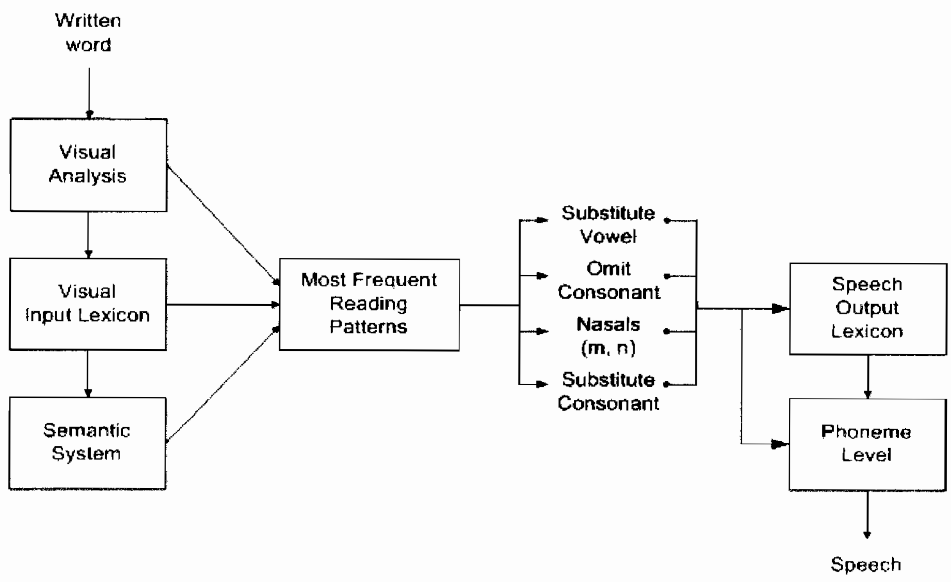


Figure 3.4. A reading model that models the most frequent reading and spelling errors as a result of dyslexic children reading aloud controlled vocabulary in BM.

Referring to the model in Figure 3.4, the most frequent reading patterns are included to show that when dyslexic children read aloud single words in BM, six possible alternative routes are used in which the reading produce these errors. The routes are as follows:

- The first route is given by 'visual analysis' → 'visual input lexicon' → 'semantic system' → 'reading pattern' (can be either one or combination of the errors) → 'speech output lexicon' → 'phoneme level' → speech.
- The second route follows that of 'visual analysis' → 'visual input lexicon' → 'reading pattern' (can be either one or combination of the errors) → 'speech output lexicon' → 'phoneme level' → speech.
- The third route takes the path of 'visual analysis' → 'reading pattern' (can be either one or combination of the errors) → 'speech output lexicon' → 'phoneme level' → speech.

- Or alternatively follows either three of the routes mentioned, skipping the 'speech output lexicon' process. This normally happens when one recognize simple, familiar words, such as *aku* or *ibu*.

The most frequent reading patterns are introduced at the point where the child is about to trigger his/her speech module and phoneme module before a pronunciation is produced. When the child sees a word, his/her cognitive processes trigger the modules necessary to process the word using either routes as mentioned. The error patterns can be regarded as outputs from either the 'visual analysis' → 'visual input lexicon' → 'semantic system' or the 'visual analysis' → 'visual input lexicon' or the 'visual analysis' module by directly accessing the knowledge of word pronunciation. This means that the errors are produced after processing the word but before pronunciation system is accessed (accessing knowledge of how to pronounce a word and the phoneme storage before articulation in invoked).

The introduction of the reading and spelling error patterns into the model is seen important to show when the errors might occur in the cognitive processes and what are the most frequent errors made when a dyslexic reads single words. These errors, although obtained from BM reading of single words, are general to extend (excluding 'nasals' that are not apparent for English) and can be used for English single word reading too as the errors obtained replicate that of Sawyer et al. (1999). Hence, the model can represents word reading with errors for BM and English.

Illuminating the errors in the reading model can serve as a useful guidance for the development of an ASR model specifically for dyslexic children. Why? Simply because, for ASR to generate a more accurate recognition it needs to be fed with the words' pronunciation models that also include these errors.

3.2.4 Improvement of the ASR model

In accordance to the pronunciation models as discussed in Section 3.2.2, the spelling and reading errors or miscues, as termed by automated reading tutor researchers, are also incorporated into the lexicon. The lexicon serves as a fundamental element to be embedded in the ASR component to generate an ASR model. Every non-words emerged from the errors made are also included in the lexicon. The lexicon is regarded as the essential component of the ASR model because without such lexicon, the ASR model cannot exactly model the reading pattern of dyslexic children in BM. The active lexicon tells the model of what to train and build its knowledge on. The lexicon is represented, in Worldbet, by the pronunciation models of the chosen words and their miscues as such:

```
aku = A kh U;        ade = A d e;
aki = A kh I;        pada = ph A d A;
apa = A ph A;        pade = ph A d e;
ape = A ph e;        baca =  bc b A ts A;
 ada = A d A;        bica = bc b I ts A; …
```

The above are some examples of the pronunciation of read words and target words in Worldbet. Note that, all of the miscues are of 'substitute vowel' category where for example, the vowel 'u' in *aku* is replaced by 'i' producing an incorrectly read word *aki.* These pronunciation models regard the errors as active lexicon and are included in the ASR model as an important element. The representation of the lexicon in such a way is as suggested by related studies as discussed in Chapter 2, Section 2.5.5.

To illustrate the connection of the reading model and the ASR model, Figure 3.5 depicts the links of these two models. The lexical component is highlighted as the key component that connects the reading model (which includes the most frequent error patterns as the results of the connection failure of the cognitive processes) and the ASR model.

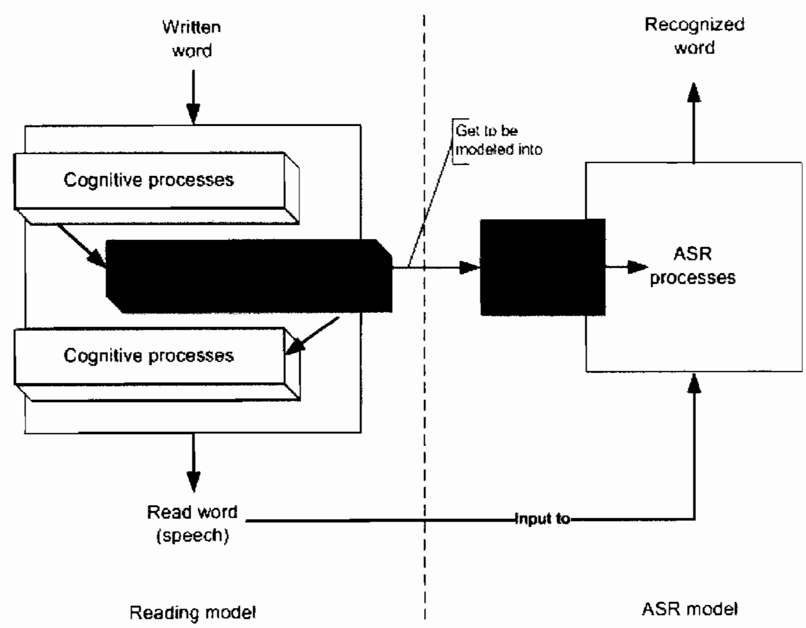


Figure 3.5. The reading model and how it supports the ASR model.

Referring to Figure 3.5, the connection substantiates two factors, in particular:

- the speech (read word), which is the outcome or output of the cognitive processes of dyslexic reading becomes the input to the ASR model. Note that the cognitive processes illustrated in Figure 3.5 represent the specific cognitive processes such as ‘visual analysis’, ‘visual input lexicon’, ‘semantic system’, speech output lexicon’, and ‘phoneme level’. Hence, whatever happens during reading (in the cognitive processes), if it is incorrect, the speech bears the error(s) in it that shall challenge the ASR to recognize it. Why? Because recognizing phonetically similar words is challenging for ASR.

- with the knowledge of the most frequent errors, it is easier to construct a more dyslexic-tuned ASR model that can produce better recognition, focusing on the lexical model. Therefore, the most frequent errors serve as the fundamental element to be included in the ASR model for dyslexic children. The justification is that, the inclusion lead to a better recognition performance and a better error detection of the ASR recognizer.

The ASR model proposed, as depicted in Figure 3.6, is enhanced from the basic hybrid architecture as illustrated in Figure 2.8. The improvement is concentrated on two parts. The first part concerns with the lexical model that includes the most frequent errors and phoneme refinement. The second part is the algorithm for miscue detection.

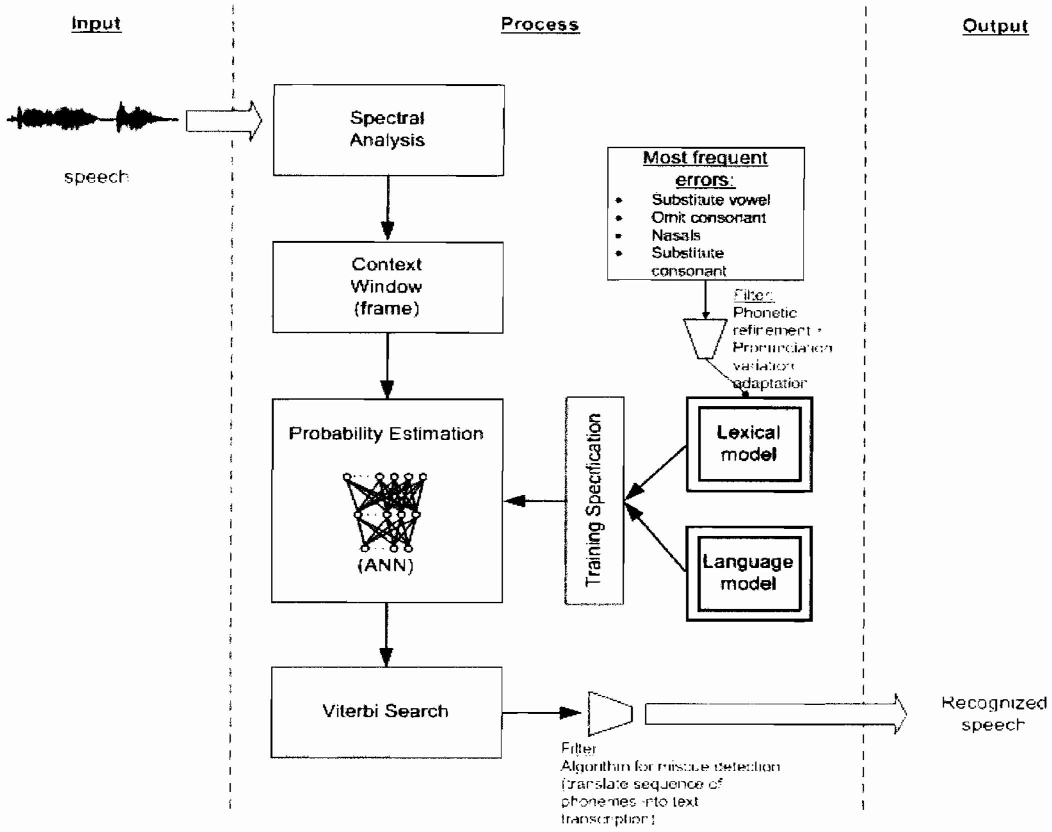


Figure 3.6. The proposed ASR model with inclusion of reading errors as active lexicon and phoneme refinement.

First, the **lexical model** represents the vocabulary that contains pronunciation models which includes the pronunciation models of miscues that fall under the most frequent error patterns obtained. Usually for speech recognition, vocabulary only contains the correct pronunciation models for target words. But for an ASR model specifically targeting on **dyslexic children's reading**, the errors are important to provide information for the ASR to train on. Since the vocabulary contains phonetically similar words and recognition accuracy is the ultimate target, developing an ASR system always faces with issues concerning phonetic recognition. As dyslexics read with many errors of phonetic substitutions (such as vowel and consonant substitutions, omissions, and reversals of letters), introducing the pronunciation models of miscues together with target words' pronunciation models is promising. Thus, incorporating them into the active lexicon with phonetic refinement consideration is nevertheless an important aspect to help increase the recognition accuracy. The **phonetic refinement** and pronunciation adaptation filter, which act as a rule for the conversion of the miscues into the lexical model is thus regarded as means to represent the pronunciation model of each word and related miscues. According to this filter, the words/miscues are modeled so that they conform to the phonetic refinement and be regarded as adaptation to the actual pronunciations of target words. The phonetic refinement and adaptation filter is introduced in order for the ASR to achieve better recognition accuracy (lower WER) when dealing with dyslexic children's reading.

Secondly, the miscue detection filter that concerns with an algorithm that takes the most probable **phoneme sequence** resulted from the Viterbi search, translates the phoneme sequence into its corresponding text, and compares that to the target word. A miscue is detected if the transcribed text fails to match the target. This filter is introduced in the model to enable miscue detection and aid the ASR's miscue detection ability for higher MDR.

Again, it is emphasized that although this ASR model is improved in consideration of dyslexic children reading single words in BM, it could also be generalized to English single words by populating the active lexicon with pronunciation models of English words. The recognition could still works because it is based on the most frequent errors obtained, which is similar to English's errors as discussed in Chapter 2. In addition, from observation, all ten participants' readings of isolated words do include the most frequent errors as mentioned. Hence, the inclusion of the most frequent errors generalizes the model to cater for more dyslexic children.

3.3 Task 3: ASR Recognizer Development

In this task, a recognizer of the ASR model is developed. The iterations of recognizer's network training are continued until optimum percentile of speech recognition accuracy with the best WER is achieved. The training is performed using HMM/ANN hybrid method as reviewed.

The hybrid between HMM and ANN makes a powerful method for speech recognition as discussed. ANN with its powerful classification ability is used to train and get the best probability estimates of categories being trained. This then serve as the input to HMM for evaluation. Since CSLU Toolkit is the only tool that deploys HMM/ANN method and the only one available for free (for research purposes), the justification of using this tool for this research is then justified. In addition, CSLU toolkit has been trained with children's speech which gives a significant importance to this study.

To use the toolkit, certain files that need to be created are as follow:

1. Create a corpora file that contains all the selected words from the vocabulary, i.e. the words that fall under the most frequent error patterns.
2. Create an info file that provide information to the toolkit on how to find examples for training, development, and testing and how to select the data for required partition such as filtering using the format specified by the toolkit.
3. Create a lexicon file (lexical model), which includes pronunciation models of words to be trained and tested by the hybrid network.
4. Incorporate words that fall under the most frequent errors as pronunciation adaptation into the lexical model with as the phonetic refinement.

5. Create a grammar file, which specifies the 'grammar' of the read word, i.e. the order of which the words are being spoken or read. For isolated words, the grammar contains specification of 'silence > word > silence'.
6. Create a part file, which specifies the number of parts that the phonemes are split into and what context clusters to be used. Context clusters refer to the context-dependent definitions used for the phonemes.
7. Use the files and train the hybrid network and perform testing. Stop when optimum recognition has been achieved.

The steps are illustrated in Figure 3.7. The "populate lexicon" process encompasses modeling the lexicon with regards to phonetic refinement and pronunciation variation adaptation of the corpus, as proposed in the ASR model (see Figure 3.6). Note that training can be performed more than once in order to obtain optimum recognition accuracy. For training, only the training info file (`train.info`) is used that describes the dataset used for training. Once trained, the network is ready to be tested to measure how good its recognition performance is in terms of accuracy, measured in WER. For testing the trained network, development dataset is fed and accuracy is obtained. In order to obtain optimum accuracy, another cycle of training can be performed. The process iterates until optimum accuracy is achieved on the development dataset and only then it is tested on the test dataset to evaluate the final network. The final network is regarded as the ASR recognizer with the highest recognition accuracy on both development and test datasets that can be used for further evaluation. The evaluation includes measuring MDR and FAR of the ASR recognizer.

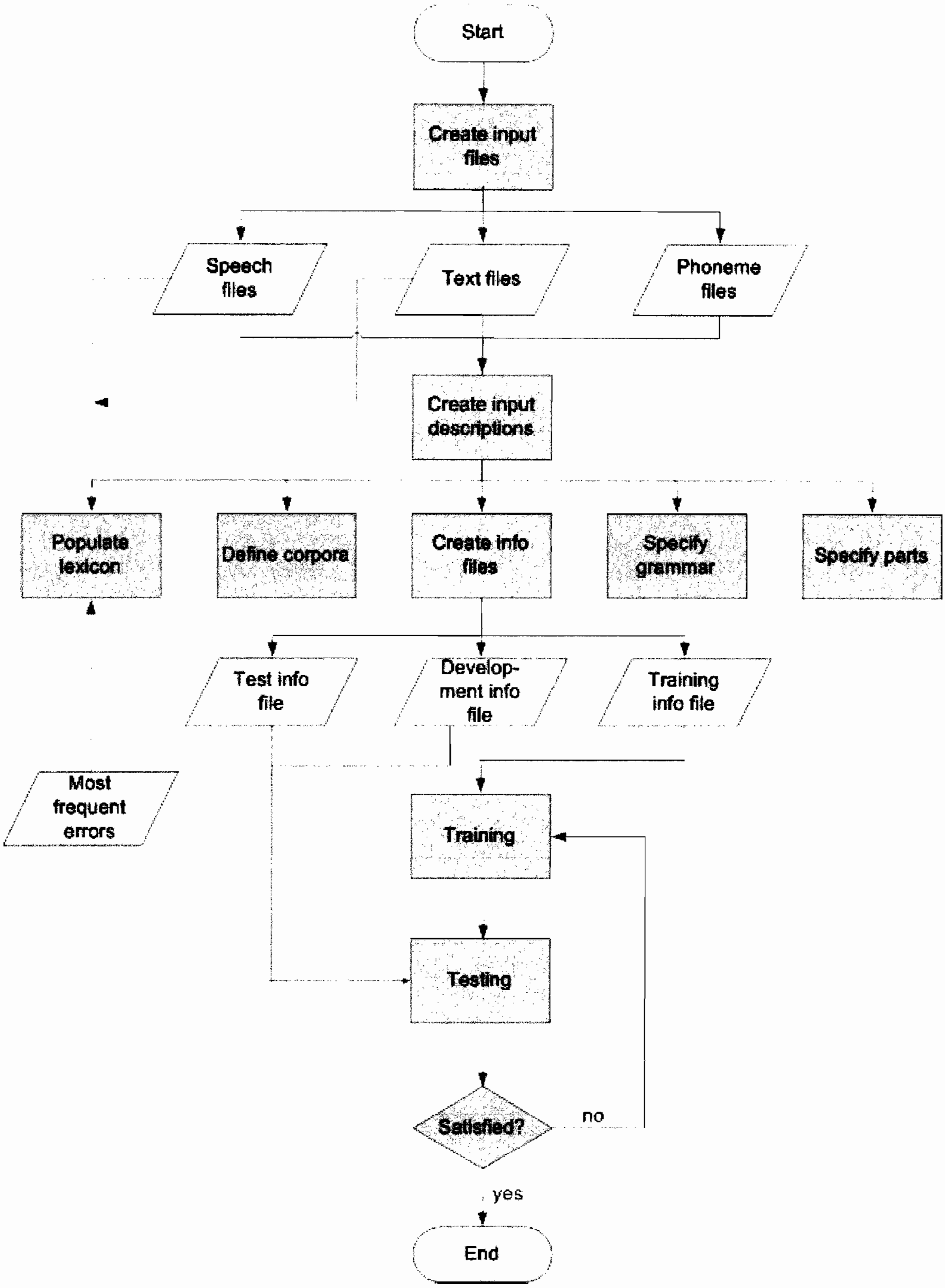


Figure 3.7. The process flow of ASR recognizer development.

3.4 Task 4: Evaluation

The recognizer serves as means toward evaluating the proposed model. Evaluation is performed to measure the performance of the recognizer in terms of its accuracy, which is based on the ASR model. Accuracy in this respect refers to the accuracy of the recognizer to recognize the read words and the accuracy of detecting reading miscues. The evaluation measures the performance of the proposed recognizer using three metrics – WER, MDR, and FAR.

3.4.1 The WER evaluation

Measuring the accuracy of a recognizer using WER enables it to be measured in terms of its ability to perform recognition with less word error. Thus, the WER evaluation aims at achieving the lowest word error. This is performed by feeding the test dataset into the ASR recognizer that later outputs the recognition accuracy in recognition rate and WER (WER = 100% – recognition rate). The test dataset are automatically selected by executing a CSLU toolkit's built-in *tcl* command (`find_files.tcl`) as described and illustrated in Chapter 4. Figure 3.8 illustrates the process to obtain the WER.

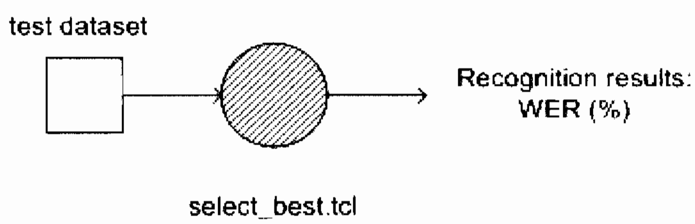


Figure 3.8. Running `select_best.tcl` to obtain WER on test dataset.

The tcl command ensures no data overlapping as the selection is based on the info file for testing i.e. `test.info`. The info file specifies the requirements on selecting the speech files and transcription files needed to perform the WER evaluation. The requirements include the corpus name, lexicon file, type of data (speech/phoneme/text), and partition of data.

3.4.2 MDR and alignment technique

For MDR and FAR evaluation purposes, a total of 100 speech files were randomly selected among available speech files (that were not used in the training, development, and testing datasets) and used to measure its miscue detection performance. Note that the selected files are of acceptable quality where they contain less background noise and garbage. Speech files with garbage (extra words/non words spoken and recorded) or/and with significant background noise were not selected. This is to ensure the recognition accuracy is not affected. The speech files are **NOT** being used during training to ensure the integrity of the evaluation process. This way, the recognizer's hybrid network has not learned the speech patterns and thus is made equal to using real-time speech. The speech files are regarded as the miscue detection dataset where 52 speech files contain reading miscues and 48 of them are correct readings.

As defined in Chapter 2, MDR is the rate of the detected miscues over the total miscues made (Banerjee et al., 2003a). To obtain the number of miscues detected, a technique called *alignment* is used. The alignment technique is conducted according to the following steps:

1. The read word is recognized by using some of the recorded speech files, which have **not** been used in training, to serve as input to the recognizer.
2. Simultaneously, transcribe the speech files. This shall later be used as a transcript to compare the recognition result with.
3. Manually perform the alignment technique to measure the MDR and FAR (for example, see Table 3.8).

For the purpose of measuring MDR, three tokens need to be taken into consideration. *Target* tokens are the actual words that the children are supposed to read given by the vocabulary, *transcript* tokens are the words read by the children as transcribed by human transcriber, and *hypothesis* tokens are words as recognized by ASR. To illustrate the alignment technique, all three tokens are aligned and compared as described in Table 3.8.

Table 3.8. Alignment of target, transcript, and hypothesis tokens for miscue detection.

Tokens	**Word**	**Word**	**Word**
Transcript	*apa*	*tempat*	*abang*
Target	***bapa***	***pantai***	***abang***
Hypothesis	*apa*	*pantai*	*adang*
	Miscue detected	**Miscue undetected**	**False alarm**

Referring to Table 3.8, a miscue is detected for target word *bapa* because the hypothesis (*apa*) is the same as the transcript (*apa*). As for *pantai*, the miscue is undetected because the recognizer fails to detect the miscue. Since the generated hypothesis is *pantai* but the transcript gives *tempat*, the miscue is regarded as undetected. As for the third example, the hypothesis generated is *adang*, which is a miscue. However, the transcript illustrates a correct reading of *abang*. Thus, the condition is considered as false alarm. However, false alarm and methods to handle them are beyond the scope of this study as the objective is to obtain significant WER and MDR with considerably acceptable FAR as reviewed in Chapter 2. The three examples illustrate that, the comparison is made between the target and the hypothesis as well as the transcript tokens (see speech space and text space in Chapter 2, Section 2.6). The reader substitutes the word *pantai* as *tempat* when reading and clearly, the ASR failed to generate a correct hypothesis for the substitution error made.

3.5 Summary

The research framework consists of four tasks – data collection, model creation, recognizer's development, and evaluation. Each of the phases is performed to achieve the objectives as specified in Chapter 1, Section 1.4. To achieve the first objective, which is to collect suitable BM vocabulary, data collection must be performed. Methods of the primary data collection are presented in order to obtain the speech of dyslexic children while reading aloud BM single words. The data are vital for the discovery of reading patterns of the children in BM context. To achieve the second objective, methods to recognize and classified dyslexic children's reading mistakes are presented.

This then leads to the potential adaptation of the most frequent error patterns as deliverable of this task into an existing reading model of word recognition. Thus, to achieve the fourth objective, an improved ASR model is proposed using methods described in task two. For achieving objective five and six, task three and task four outline and discuss the required methods. In addition, for the purpose of achieving task four – recognizer development, a separate chapter (Chapter 5) is exclusively made to clearly and structurally describe, in details, the steps and activities required. Thus, this chapter outlines and discusses methods to achieve the objectives as specified. The next chapter discusses the model creation by presenting the results of task one, which serve as the key elements needed to improve both the reading model and the ASR model.

CHAPTER 4

THE DEVELOPMENT OF THE ASR RECOGNIZER

4.0 Introduction

In the previous chapter, methods to achieve the specified objectives (see Chapter 1 Section 1.4) have been outlined and discussed in details. The discussion also includes the results of the four most frequent error patterns that lead to proposing both the reading model and the ASR model for dyslexic children. This chapter continues to discuss the development of an ASR recognizer, which is based on the proposed ASR model. Hence, this chapter embarks on 'task three' of the research methodology. This chapter elaborates on Figure 3.7 of Chapter 3, Section 3.3, focusing on the development and training of the recognizer.

First and foremost, important input files to the ASR recognizer need to be prepared namely, 1) the speech files of the recorded reading of the participants obtained through data collection process; 2) the reading transcription files, which contain the orthographic representation of the corresponding speech files; and 3) the phoneme files that provide time-aligned phonetic labels of each of the speech files.

With the aforementioned input files prepared for training, their descriptions need to be specified namely, 1) the corpora; 2) information of datasets to be used when training, development, and testing, i.e. the info files; and 3) the vocabulary or lexical model for which the ASR is going to be trained on.

Note that the lexical model is what connects the reading model and the ASR model. The proposed reading model suggests that the most frequent error patterns to be included for reasons discussed in Section 3.2.3. The ASR model then incorporates the most frequent error patterns in its lexicon component and employs the information to enhance the accuracy of the ASR recognizer specifically for dyslexic children. Hence, this chapter discusses the methods and process of developing an ASR recognizer based on the proposed ASR model.

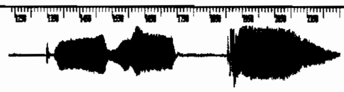
Prior to evaluation, the performance of the ASR recognizer needs to be at par with other similar recognizers that concerns with children's speech while reading. Therefore, the ASR recognizer is tested on the test dataset. This is performed as the aim of this chapter is to achieve the fifth objective, which is to develop a recognizer based on the ASR model proposed. Hence, this chapter outlines the process flow as depicted in Figure 3.7 of Chapter 3 and elaborates on each process. Section 4.1 deals with input to ASR recognizer – the speech files, the transcription files, and the phoneme files. Section 4.2 discusses the descriptions prior to training where the most frequent errors play a significant role as a filter in selecting words that contain these errors to be modeled. Section 4.3 details the training process, the ASR network, and the development and testing results while Section 4.4 discusses the results when phoneme refinement is applied. The final section, Section 4.5, summarizes the chapter.

4.1 Setting the 'Stage'

The first step towards developing a recognizer is to create the files needed prior to training. The files are the speech files gathered during data collection, the transcription files, and the phoneme label files. The text or transcription files and phoneme files for each speech file (the read word) involved has to be created prior to training. Table 4.1

illustrates the relationship of the files. Below is a brief description of each type of files needed prior to training a speech recognizer.

Table 4.1. The relationship of speech file, transcription file, and phoneme file.

File type	Example	Description
Speech file (.wav)		Contains the speech signal of *kelapa*
Transcription file (.txt)	kelapa	Contains the transcription of the speech
Phoneme file (.phn)	MillisecondsPerFrame: 1.0 END OF HEADER 1333.397827 1350.511597 kc 1350.511597 1380.039429 kh 1380.039429 1533.895142 & 1533.895142 1605.383667 l 1605.383667 1723.495117 A 1723.495117 1903.770630 pc 1903.770630 1934.852539 ph 1934.852539 2254.996826 A	Contains the time-aligned phonetic labels of the speech

4.1.1 The speech files (.wav)

The speech files are obtained during data collection. These files contain read speech of dyslexic children who participated in the study. Read speech is the speech recorded when the children read aloud a single word in BM using SpeechView. Figure 4.1 illustrates examples of three speech files for the words *baca*, *bunga*, and *pada* respectively. The shaded section of each speech file highlights and magnifies the selected speech of the .wav files.

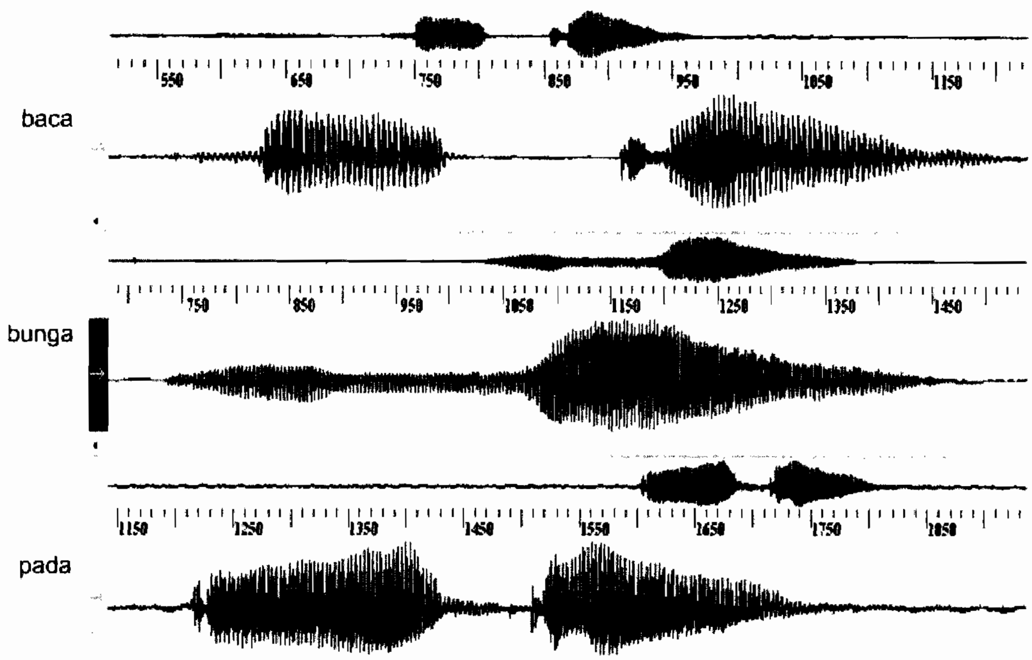


Figure 4.1. Example of speech files recorded displayed in SpeechView.

These files are used to train the ASR network before a recognizer can be created. Note that, as mentioned in Chapter 3, only speech files of which the spoken attributes correspond to the most frequent error patterns are being used and trained. The most frequent error patterns are vowel substitution, consonant omission, nasals, and consonant substitution.

4.1.2 The transcription files (.txt)

The transcription files are created manually. A transcription file contains the orthographical form of a word as read by an individual participant. This means that the speech is transcribed into its corresponding spelling.

For a speech file of the word *kelapa*, a transcription file is created which contains its spelling as *kelapa* and the file is saved as DC-X.wordY.txt, where X represents the ID of a participant whom the speech file represents and Y for which session that the speech is recorded. Therefore, for this particular word *kelapa* spoken by dyslexic children ID 1 or DC1 in session 1 of reading recording, the file is named DC-1.kelapa1.txt. The file naming convention is used to make sure that each file is unique and can be easily differentiated when running training and testing. The transcription files are validated by re-listening to the speech file and its transcription is re-checked.

4.1.3 The phoneme files (.phn)

Referring to the example as presented previously, a time-aligned phoneme file needs to be created that specifies the position of every phoneme in *kelapa* spoken by DC1. The phoneme file can also be called the label file where every phoneme is hand-labeled with its corresponding sound symbol. An example of a phoneme file of *kelapa* is as depicted in Figure 4.2 below.

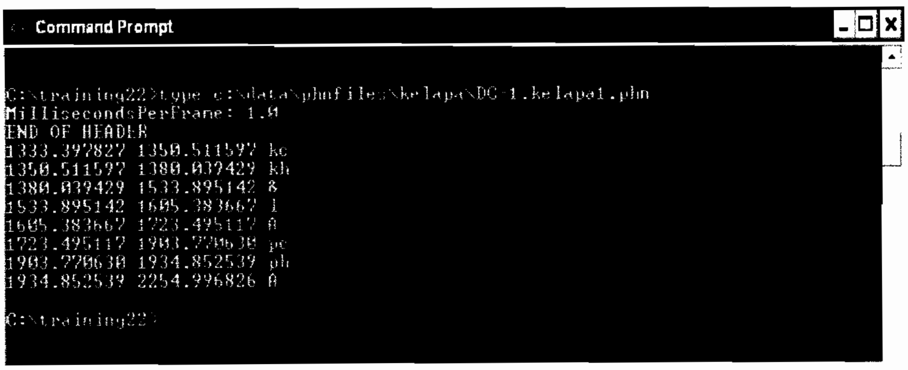


Figure 4.2. The phonetic labels (in Worldbet) of every phoneme of the word *kelapa* and their positions given in millisecond (ms).

In Figure 4.2, the read speech of the word *kelapa* is labeled with its corresponding phonemes that make up the word. The kc kh symbols represent the sound of the letter 'k', the & symbol represents the sound that the letter 'e' makes, and so on. The phoneme file also provides the information of the phoneme position in a speech file in millisecond. Referring to Figure 4.2, the position of the phoneme kh in the speech is at 1350.511 ms to 1380.039 ms. The phoneme files are created manually in SpeechView as illustrated in Figure 4.3, by manually labeling the phonemes at their target positions determined. Once completed, the label is saved as <*filename*>.phn for phoneme file. Figure 4.3 is actually an alternative view to Figure 4.2 where it can be clearly seen that the speech is segmented and labeled accordingly with the corresponding phonemes in each segment of the speech file.

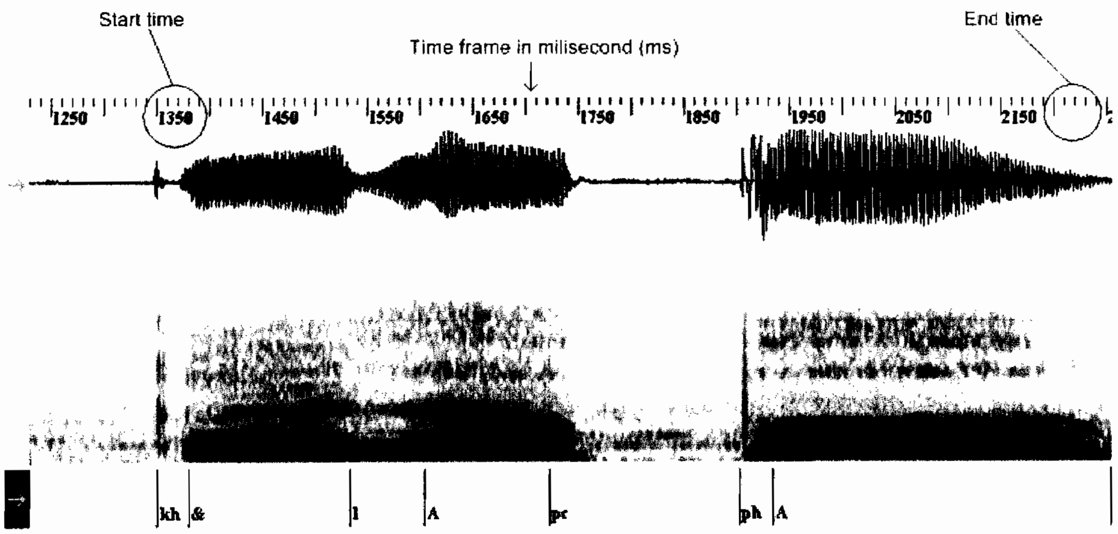


Figure 4.3. The view of speech window (top), 2-D spectrogram window (middle), and the label window (bottom).

4.2 Creating Descriptions Files

The previous section discusses the speech, phoneme, and transcription files that need to be created prior to training. In addition, this section continues to discuss the input descriptions that should be created before training could be performed. The input descriptions are necessary to provide fundamental information to the HMM/ANN network before training can be performed. The input descriptions are prepared in five separate files namely a lexicon file that specifies the lexical model, a corpora file, a grammar file, a parts file, and an info file for each training, development, and testing datasets.

4.2.1 The lexical model as the key component in ASR

The lexical model is the focus of this study where the most frequent reading errors obtained during data collection and analysis phase are modeled into the ASR model as discussed in Chapter 3, Section 3.2.4. To train the recognizer, the proposed lexical model is used. The language model or the grammar constitutes to the speech files is straightforward since the ASR recognizer is dealing with isolated words. Hence, what makes a significant impact towards the performance of the ASR recognizer is thus focused onto the lexical model.

Despite having to specify text transcriptions and phoneme files prior to training, the lexical model, which is the key component of an ASR recognizer that lists all the words (including mispronunciations) is constructed for the ASR to train on. A snippet of the lexicon is shown in Figure 4.4. The lexical model models the vocabulary of which the ASR is trained on and tested. In this case, the vocabulary consists of all target words and their mispronunciations that fall into the four most frequent errors obtained – "substitute vowel", "omit consonant", "nasals", and "substitute consonant".

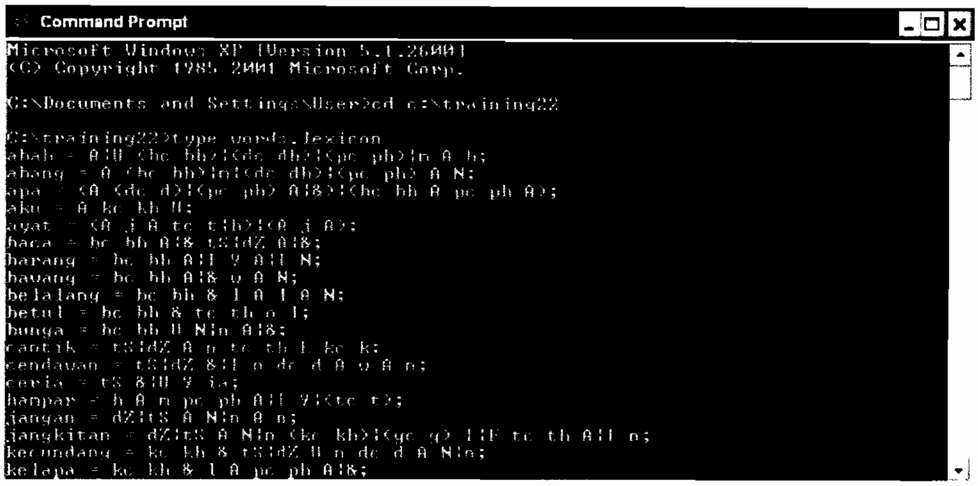


Figure 4.4. A snippet of the lexicon file containing target words' pronunciations as well as mispronunciations represented in Worldbet, separated by the OR (|) operator.

The inclusion of mispronunciations into the lexical model for training the ASR is seen important, **especially** when dealing with the recognition of dyslexics' readings. Due to their difficulties, they produce many mistakes or miscues even when reading simple, familiar words such as *wad*, producing errors like "wed", *wap*, and *wan*. Table 4.2 depicts examples of mispronunciations of the word *wad* and their percentage of occurrences (see Appendix C for a list of target words, which are modeled, and their corresponding mispronunciations and percentage of occurrences).

Table 4.2. The mispronunciations of wad and their frequencies that later suggest future potential feedback.

Word	Correct / Miscues	Frequencies (%)	Potential feedback
wad	*wad*	36.36	Praise (correct reading)
	wan *wap* "yad" "wed"	23.64 20.00 3.64 1.82	Trigger feedback, e.g. TTS component to utter the correct pronunciation

The miscue modeling presented in the lexical model manages to increase the recognition accuracy because the recognizer has learnt the miscues. Other than that, it would also ease reading tracking process since the recognizer has the knowledge of what common errors to be expected when a dyslexic child tries to read the word *wad*, for example. Reading tracking is an important feature in any automated reading tutor, as described briefly in Chapter 1, Section 1.3. However, the scope limits this study to focus on the ASR recognizer that is built with optimum accuracy by carefully modeling its lexical model. Thus, it is important to be able to recognize miscue in order to trigger suitable feedback or correction.

The lexicon could also look like the following with the inclusion of '_miscues' at the end of each mispronunciation as it allows distinguishing between correct and incorrect pronunciations of a word for example:

Example 1:

```
bapa        = bc b A pc ph A ;

apa_miscue  = A pc ph A ;
```

Example 2:

```
ada          = A dc d A ;
dapa_miscue = dc d A pc ph A ;
```

The problem of modeling miscues this way is that there will be a clash should the error of a target word matches another target word with the same grapheme representation. For example, consider that a target word *apa* is also present in the lexicon: `apa = A pc ph A`. Notice that the phoneme sequence are identical to `apa_miscue` as presented in Example 1. If the Viterbi search returns the most probable phoneme sequence for the speech the ASR is trying to recognize, which target word should it be assigned to? It could be *apa* or a miscue for the word *bapa*.

Alternatively, they may not even need to be labeled as miscues. The postfix '_miscue' can be removed from the word and when the recognizer recognizes *apa*, if it is not the expected word *bapa*, it shall identify the recognized word as an error. Even though *apa* is also a valid word in BM, it is treated as an error because the target word that one is suppose to read is *bapa*. Therefore, when someone read *bapa* as *apa*, it is clearly a reading mistake and should be treated so by a recognizer that is meant to recognize reading. This way, feedback to correct the reading could be triggered. However, the feedback triggering mechanism is beyond the scope of the study. Nevertheless it is important that the recognizer is able to detect reading mistakes to provide a more accurate intervention.

The miscues are normally accepted and treated as active lexicon in the lexical model (Fairweather et al., n.d.; Mostow et al., 1994; Williams et al., 2000; Mostow et al., 2002; Banerjee, Mostow, Beck, & Tam, 2003; Banerjee, Beck, & Mostow, 2003). This means that the miscues are included as an active vocabulary for ASR. For example, consider the miscues for target word *ayat*: *ayah* and *aya*. Each of these is modeled as an individual active lexicon as:

```
ayat = A j A tc t;

ayah = A j A h;

aya = A j A;
```

For better performance this study proposes the miscues to be modeled as alternative pronunciations of the target words. Alternative pronunciation allows pronunciation variations to be included, which means that the miscues are regarded as the variants to the pronunciation of the target words. When defining the pronunciation of a particular target word the alternative pronunciations are included into the active lexicon by separating them using the '|' (OR) operator, as illustrated in Figure 4.5. Consider the aforementioned example that gives *ayah* (substitution of 't' with 'h') and *aya* (omission of 't') as miscues for *ayat*. Hence, in the proposed lexical model, the miscues are modeled as alternative pronunciations as such:

```
ayat = (A j A t|h) | (A j A) ;
```

For illustration, consider the word *ayat*, which is incorrectly read as *ayah*. When the recognizer recognizes its phoneme sequence as `A j A h`, the phoneme sequence is translated into its corresponding graphemes that spelled *ayah* and the spelling is compared with the target word *ayat*. This can be regarded as a filter that filters whether or not a recognized word (hypothesis) is a miscue. Since 'h' is not equal to 't', *ayah* is regarded and treated as a miscue even though it is a valid word in BM, which semantically means "father". The translation is performed by executing the `asr_filter.tcl` command that has been enhanced to include the phoneme-to-grapheme translation algorithm for miscue detection (Appendix D). The filter translate the most probable phoneme sequence output by the Viterbi search (in ASR process of Figure 4.5) and transcribe them into text transcription. The text transcription is then compared with its target word. Any deviation from target concludes that a miscue has been detected. The filter also considers acceptable pronunciation of certain words to be regarded as correct even though its spelling differs slightly from the target, e.g. 'cantek' to *cantik*, 'betol' to *betul*. It illustrates the ability of the recognizer to recognize phonetically similar phones such as I – E and U – O and because these words are usually pronounced as such and NOT according to its corresponding phones associated with the letters, the pronunciations are accepted as correct reading.

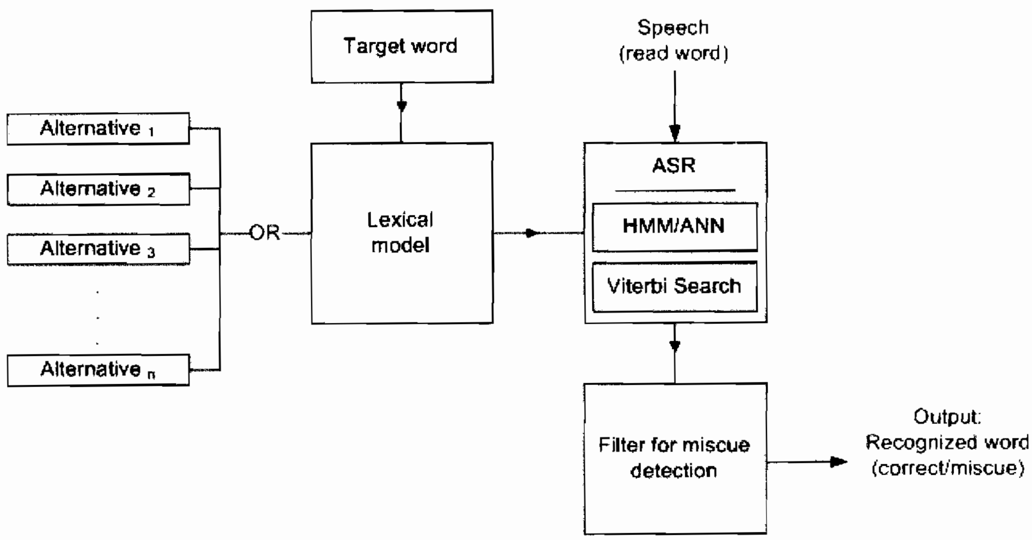


Figure 4.5. Alternative pronunciations of target words modeled into the lexical model.

In conclusion, modeling the miscues this way helps to increase recognition accuracy (see the results in Chapter 5). It allows the miscues to be regarded as pronunciation variations and reduce the number of phonetically similar active lexicon that could lead to duplicates or ambiguous representation of target words and the miscues. Thus, the adaptation of variations into the lexical model as discussed enriches the model to be more robust towards dyslexic children's variability (due to their difficulties) when reading the words.

4.2.2 The corpora

The corpora file provides basic corpus information such as the location of speech files, phoneme files, and transcription files. The filename format is also specified and it is best to name all the speech files, phoneme files, and transcription files according to the specified format. The filename format (in 'format' field) is specified in a regular expression written in *tcl* as shown in the example of corpora specification below.

```
corpus:    bmwords
           wav_path   /data/speechfiles
           txt_path   /data/txtfiles
           phn_path   /data/phnfiles
           format     {DC-([0-9]+)\.[A-Za-z0-9_]+}
           wav_ext    wav
           txt_ext    txt
           phn_ext    phn
           cat_ext    cat
           ID:        {regexp $format $filename filematch ID}
```

The `ID` field is also a regular expression to determine the speaker ID associated with a speech file, phoneme file, or text file. The filenames are case sensitive therefore `DC-1.ayat2.txt` is not the same as `dc-1.AYAT2.txt`.

4.2.3 The grammar and parts specifications

Grammar specification specifies the grammar or structure of a speech attributes in a wave file. The grammar specification for single word reading is simpler than when dealing with continuous speech. Since the study concerns with single word reading in BM, the grammar specification is `$grammar = [*sil%%] $word [*sil%%];` where `*sil%%` represents silence and/or garbage and `$word` represents the word spoken corresponding to the speech files.

The parts specification can be referred to as the phoneme specification to determine whether a phoneme is *context-dependent* or *context-independent.* Context-dependent means that a phoneme is either dependent upon its preceding and/or following phonemes or context. On the other hand, context-independent modeling does not consider the preceding and/or following phonemes or context. Context-dependent modeling is performed to allow for variations in the speech signal for the same phoneme to be regarded and trained. To illustrate, see Figure 4.6 where *aku* is modeled in a context-dependent pronunciation model.

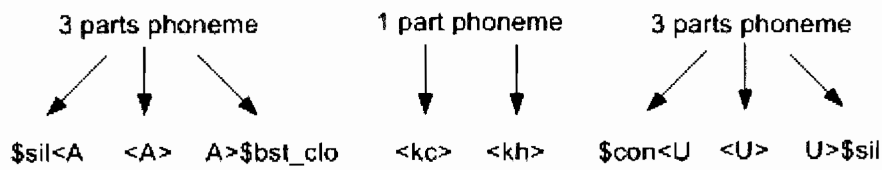


Figure 4.6. The context-dependent phoneme model of the word *aku* (source: Husniza & Zulikha, 2010).

The context-dependent modeling is performed according to the phonetics and phonological system of BM. Thus, the context clusters are defined based on BM phonetic and phonological system as described in Indirawati and Mardian (2006). The context clusters are the broad categories of context where each phoneme can be associated to. In this case, the context clusters are:

- `$fnt` for front vowels – a, e, and i;
- `$mid` for middle vowels – è;
- `$bck` for back vowels – u and o;
- `$dip` for diphthongs – ai, au, and ia;
- `$dig` for digraphs – ng and ny;
- `$con` for consonants (defined to be context-independent as they are least influenced by their neighbouring phonemes and thus always pronounced in the same manner);
- `$sem-vow` for semi-vowels – w and y;
- `$vib` for vibration – r, and
- `$bst_clo` for burst closures.

For semi-vowels 'w' and 'y' and vibrate letter 'r', they depend upon their left and right context. Other consonants, such as 'c', 'g', and 'k' are defined as context-independent (having only one part) since all consonants in BM are always pronounced in the same way. The definition of the parts that a phoneme should take depends on the wave signal for each phoneme, which is visible when the 2-D spectrogram of a phoneme is examined.

For vowels, they are modeled as having three sub-phonetic parts, which means that each phoneme is divided into three parts or depends on three parts – its left context, middle context, and right context. For example, the letter 'a', which produces the sound `A` (Worldbet) is depending upon its left, middle, and right *context*. Figure 4.7 illustrates the 2-D spectrogram difference of a phoneme `A` spoken in the word *bawang* (see shaded section of Figure 4.7). The difference, although very little, does make a significant impact towards recognition as demonstrated by Dupont, Ris, Couvreur, and Boite (2005), Tsakalidis et al. (2009), and Husniza and Zulikha (2009a, 2010).

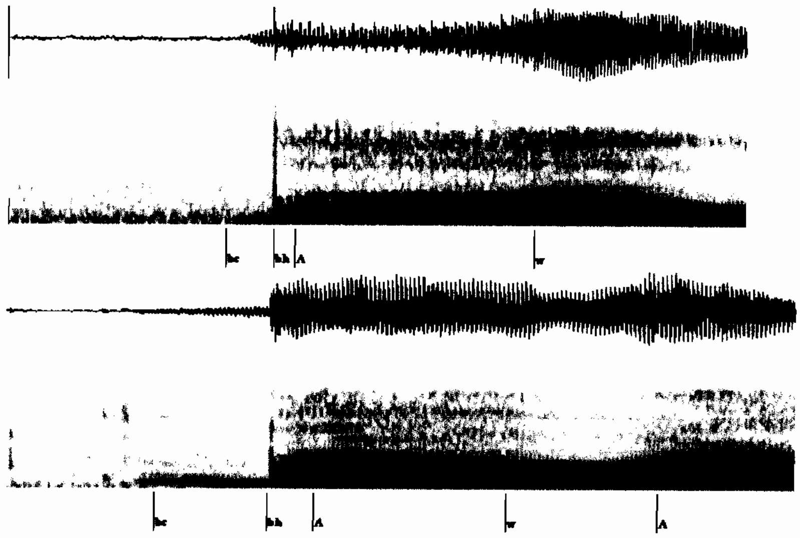


Figure 4.7. 2-D Spectrogram of phoneme A in the word *bawang*.

4.2.4 The `.info` files

The `.info` files contain information on the training, development, and testing datasets as separate files. The `.info` files describe corpus information for which the datasets will be trained, developed, and tested on. Corpus information includes the corpus name, the category path for which the categories of phonemes are stored, and the partition of data for training, development, and testing.

The speech files, transcription files, and phoneme files are divided into three datasets – training set, development set, and testing set. The division is as specified in the .info files for training, development, and testing so that it follows the 3:1:1 ratio as required by the CSLU Toolkit's. Hence, a total of 188 speech files are used for training, 53 files for development, and 48 for testing, as depicted in Figure 4.8. The files are automatically divided by running `find_files.tcl` as described later in this chapter. It is important to make sure that the data in the testing set do not occur in the training and development datasets. Thus `find_files.tcl` makes sure that the data for each dataset are unique and do not overlap based on the description that are specified in the info files.

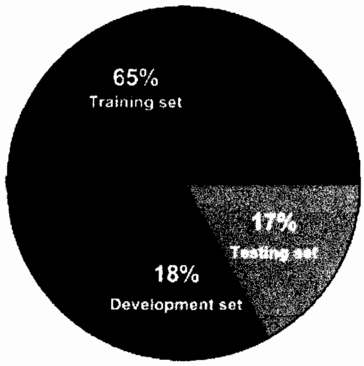


Figure 4.8. Speech files divided into three datasets according to the 3:1:1 ratio.

During training, the 3/5 dataset[5] is used for the network to learn and obtain as much as possible the general properties of the training data. Learning the general properties of the training data allows the network to be able to classify utterances which does not occur in the training data. Hence, the training dataset is prepared to train the network. Note that the development and testing datasets are not learned by the network during training. The development set is used for cross-validation where they are used to

[5] 3/5 of a total of 289 files are equal to 65% where 3/5 x 289 (total files) x 100% = 65.0519 = 65%.

(Total files = 188 + 53 + 48 files for training, development, and testing datasets respectively as mentioned in the thesis).

evaluate the network's ability to recognize phonetic categories. The final dataset, the testing set, is used to evaluate the network's performance.

4.3 The Training Process

As discuss in the previous sections, all input files and description files created are used in the training process. Without the files, a speech recognizer cannot be built and tested. The input files (the speech files and their corresponding transcription files and phoneme files) act as input to the recognizer to train on. The description files provide essential information about the corpus under training, which includes the lexical model and language model (grammar) as well as the context-dependent information (as described in the parts file). Figure 4.9 illustrates the overview of the training process.

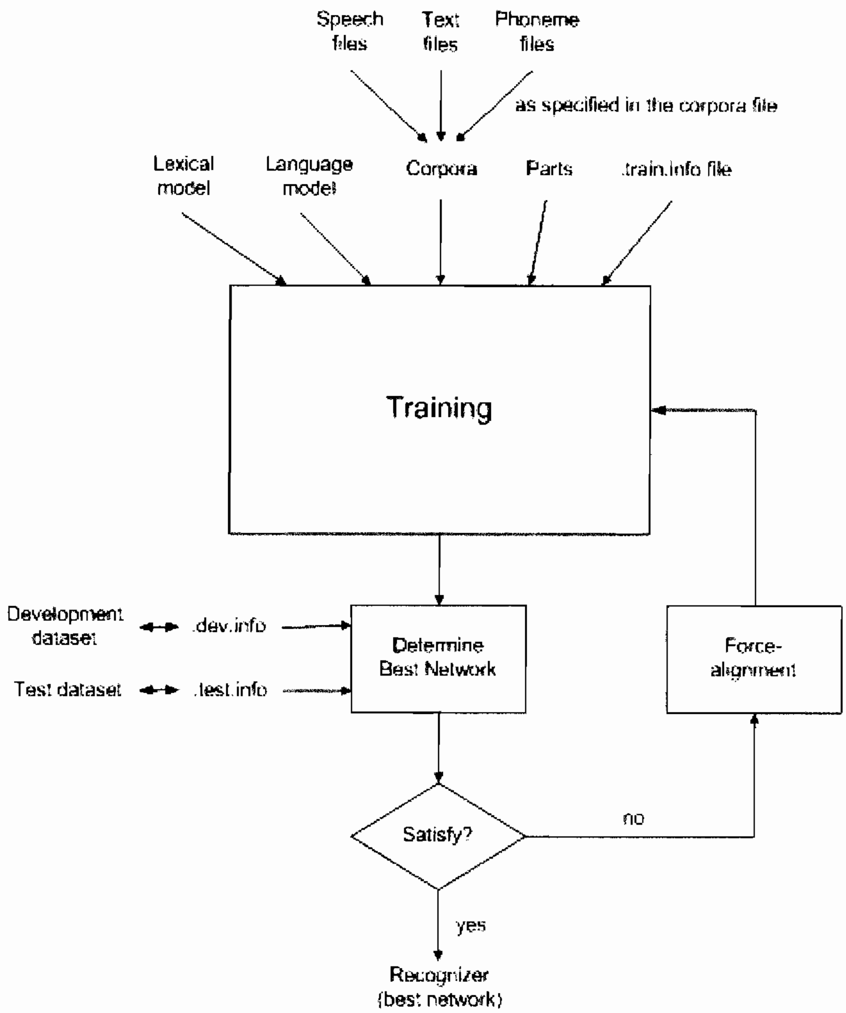


Figure 4.9. The overview of the training process for developing a speech recognizer using CSLU Toolkit.

For a complete, step-by-step instruction on how to train the hybrid network, refer to Appendix E. The instruction, adapted from Hosom et al. (2006) is used to train the network. It illustrates the steps needed to perform training from constructing the language model and lexical model to testing the recognizer built with the test dataset.

To illustrate the step-by-step process, Figure 4.10 shows the first step prior to training, which is dividing the datasets according to the ratio by running `find_files.tcl`. This command is run three times, one time for each dataset as specified in the info files for training, development, and testing.

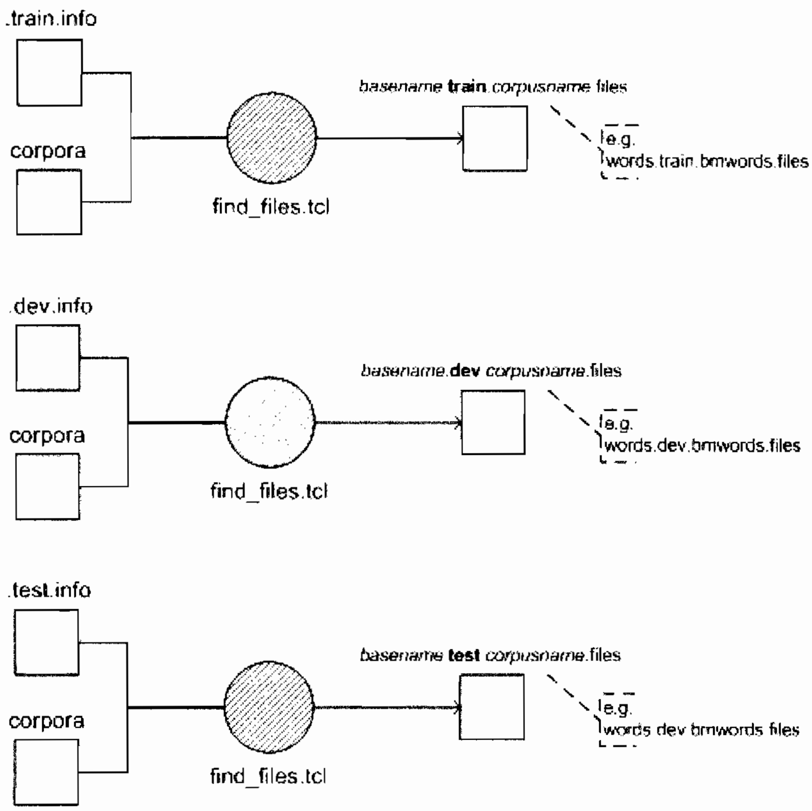


Figure 4.10. Automatic generations of training, development, and testing datasets using `find_files.tcl` command.

Once the datasets are generated, the training process begins. To train the ASR recognizer, a feed-forward, 3 layer network is used. The network is used to learn about the general properties of the training data so that classification can be performed on utterances that are not within the training dataset. An illustration of the network is depicted in Figure 4.11.

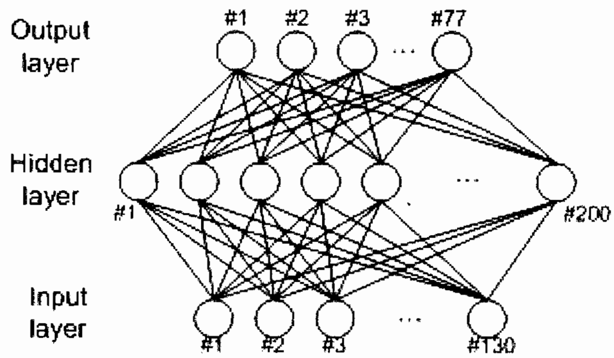


Figure 4.11. An illustration of the 3 layer feed-forward network architecture.

A number of 130 input units and 200 hidden units are employed for the standard feature of the toolkit. The number of the output units, however, depends on the total number of categories that are considered to be trained. Therefore, in this case, 77 output units are used based on the vector file created (see Figure 4.12). Noteworthy, the method chosen for training is the hybrid of HMM/ANN method for their performance as reviewed in Chapter 2.

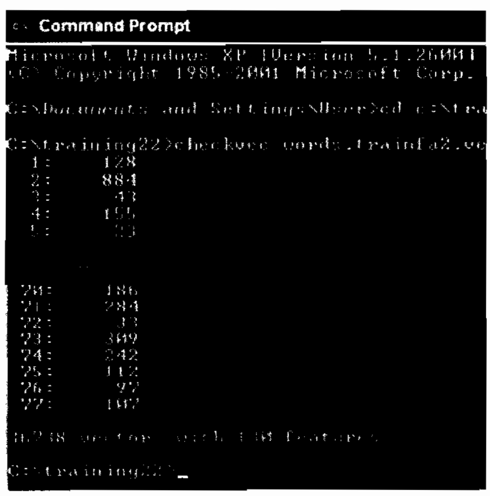


Figure 4.12. A snippet of the output of the vector file created showing the maximum number of categories on the left column, which gives 77.

The network is trained by running the `nntrain.exe` command of the toolkit generating output as depicted in Figure 4.13. The network is set to train for a total of 30 iterations using the standard features of the toolkit as suggested in the documentation. Its initial weight is set to -1.0 with learning rate set randomly at 0.05. The weight is used to multiply the signal transmitted from input layer to hidden layer while the learning rate is used to control the amount of weight adjustment in training (Fausett, 1994). The `nntrain.exe` command executes with the following parameters for the standard feature:

`-l`	allows for negative penalty (negpen)
`-sn 88 -sv88`	random seed number
`-f wordsnet`	basename of the network called 'wordsnet'
`-a 3 130 200 77 30`	architecture: 3 layers, 130 input nodes, 200 hidden nodes, 77 output nodes, and 30 iterations
`words.train.vec`	vector file

While training, the learning rate is automatically adjusted to minimize the total error. The total error should decrease in every iteration. Referring to Figure 4.13, in order to see whether or not the network is learning properly, the total error in the second iteration (41692) is divided by the total error in the first iteration (58780). In this case, the error ratio obtained is equal to 0.71, which is acceptable. According to Hosom, the acceptable ratio ranges from 0.5 to 0.9 where the ideal ratio is 0.75 (J.-P. Hosom, personal communication, March 6, 2009).

```
creating net with seed 88
negpen 0 is 0.265720
negpen 1 is 1.000000
.
.
negpen 74 is 0.232403
negpen 75 is 0.201195
negpen 76 is 0.221998
3 layers: 131 200 77
learning rate 0.050000
negative weight 1.000000
training file words.trainfa2.vec
time:13 learn_rate 0.041667; total error is 58780.757813
time:13 learn_rate 0.035714; total error is 41692.933598
time:13 learn_rate 0.031250; total error is 35380.515625
time:13 learn_rate 0.027778; total error is 31342.871094
.
.
```

Figure 4.13. The results for learn rate and total errors while training the hybrid network using `nntrain.exe` of CSLU Toolkit.

Figure 4.14 and 4.14 (cont.) depict the detailed flow illustrating the input files, the processes, and the output file(s) generated when executing a process. The output file(s) then serve as input to the following processes. The processes here represent the commands such as `gen_spec.tcl`, `gen_catfiles.tcl`, `revise_spec.tcl`, and `checkvec.exe`. Table 4.3 lists all the commands used to develop the recognizer and summarizes their functions.

Table 4.3. The commands used in training the recognizer and their functions.

Command	Function
`find_files.tcl`	To find files for training, development, and testing according to the .info files. The number of files for each dataset is determined by specifying the ratio for the three dataset. As suggested in the toolkit's documentation, training dataset consists of 3/5 of files and each development and test set comprise only 1/5 of files.
`gen_spec.tcl`	To determine the number of categories that the recognizer is going to classify, whether it is context-dependent or context-independent.
`gen_catfiles.tcl`	To use the list of files for training, created by running find_files.tcl and create time-aligned phonetic categories.
`revise_spec.tcl`	To check whether or not the number of categories to train on is sufficient. If the number of category is not enough, that particular category is tied to other categories.
`pick_examples.tcl`	To select examples of frames to train on using the output file created by gen_catfiles.tcl.
`gen_examples.tcl`	To compute acoustic features for the selected frames.
`checkvec.exe`	To check whether or not the vector file created is correct and to ensure that at least one example per category is available for training.
`nntrain.exe`	To train the neural network using the vector file created.
`select_best.tcl`	To evaluate the performance in terms of recognition accuracy of the network trained.
`fa.tcl`	To execute force-alignment.

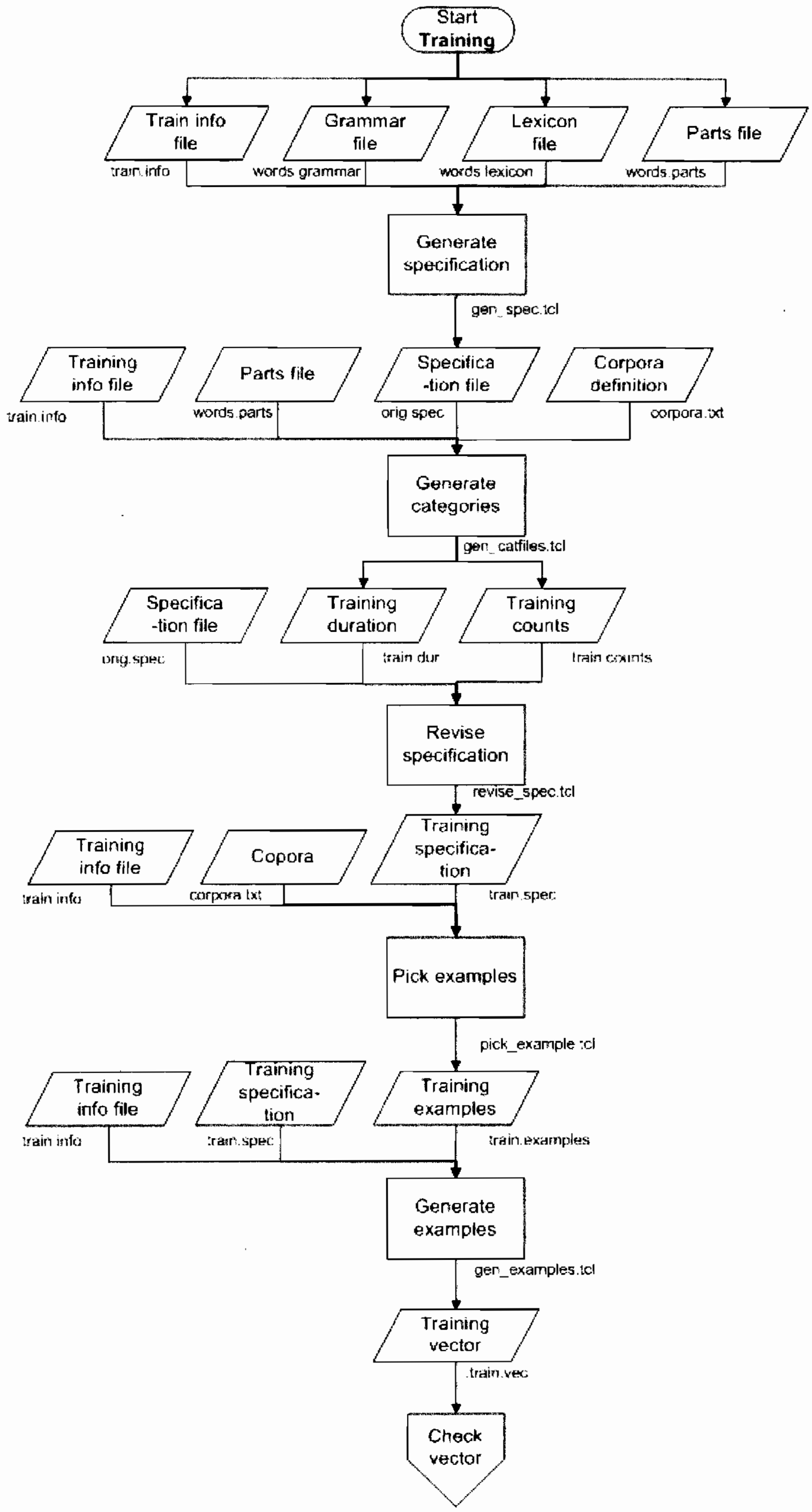


Figure 4.14. The flow of training process in detail.

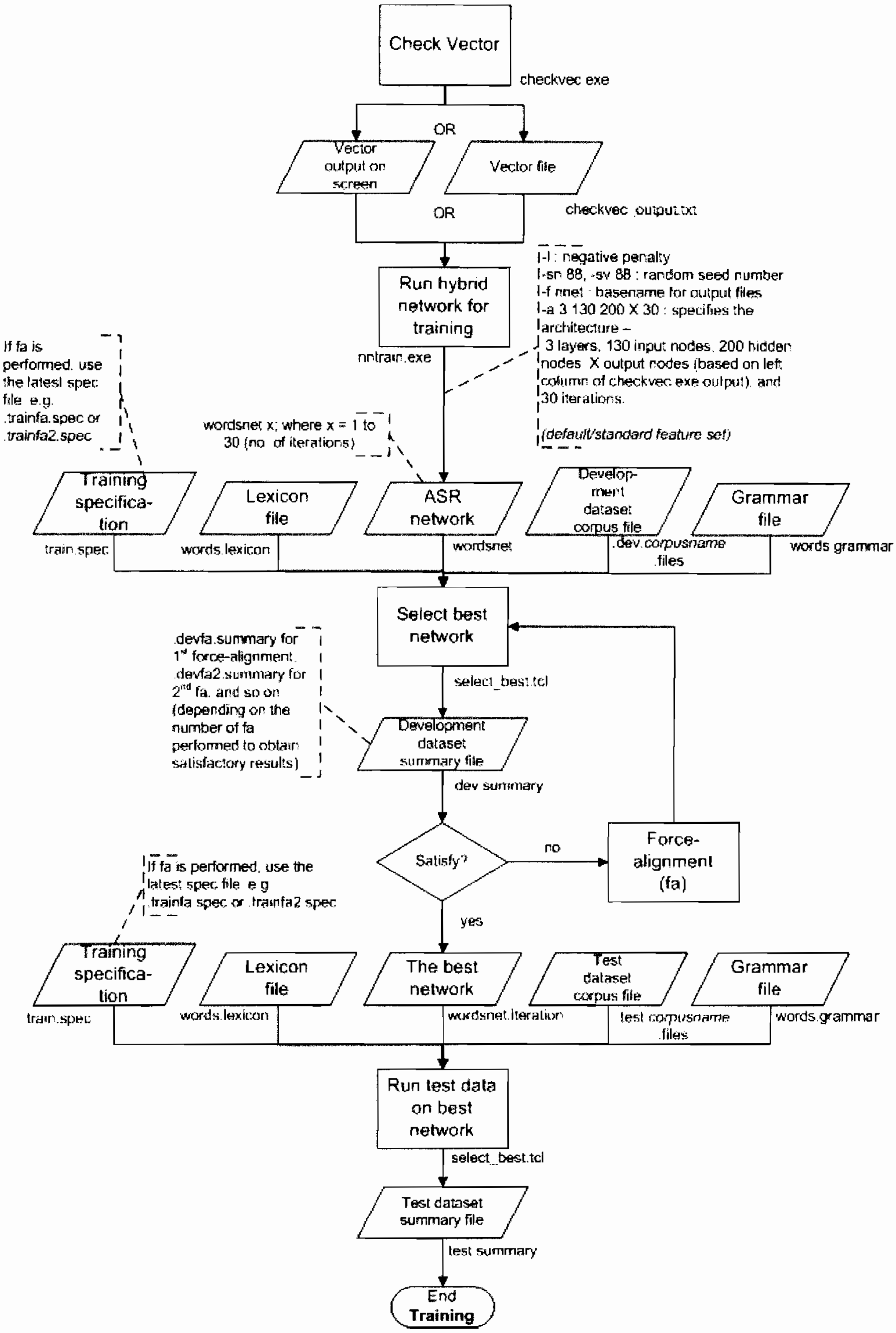


Figure 4.14 (cont.). The flow of training process in detail (continues).

Referring to Figure 4.14 (cont.), the development dataset given by `words.dev.bmwords.files` is used as input to the process `select_best.tcl` to obtain the recognition accuracy for the development set. Note that, this is not the final recognition accuracy, which should be measured using the test dataset. In this case, the `wordsnet` fails to produce optimum results. Therefore, force-alignment is performed.

Force-alignment is another cycle of training using the same dataset but by automatically generating the phoneme files. It is basically another cycle of training using the same process but with an augmented `train.info` file, which specifically specifies that the hand-labeled phoneme files are not required (in the 'require:' field). Instead, it generates the files automatically. For that purpose, `fa.tcl` code is automatically executed, which is embedded in the `trainfa.info` file. See Appendix E for details on how to create or augment the original `train.info` file for use in force-alignment process. Figure 4.15 and 4.15 (cont.) illustrate the force-alignment process, which basically repeats the same process as in training but creating forced-aligned files instead (marked by **'fa'** in the filenames). For example, the `gen_examples.tcl` creates a vector file named train**fa**.vec instead of just `train.vec`.

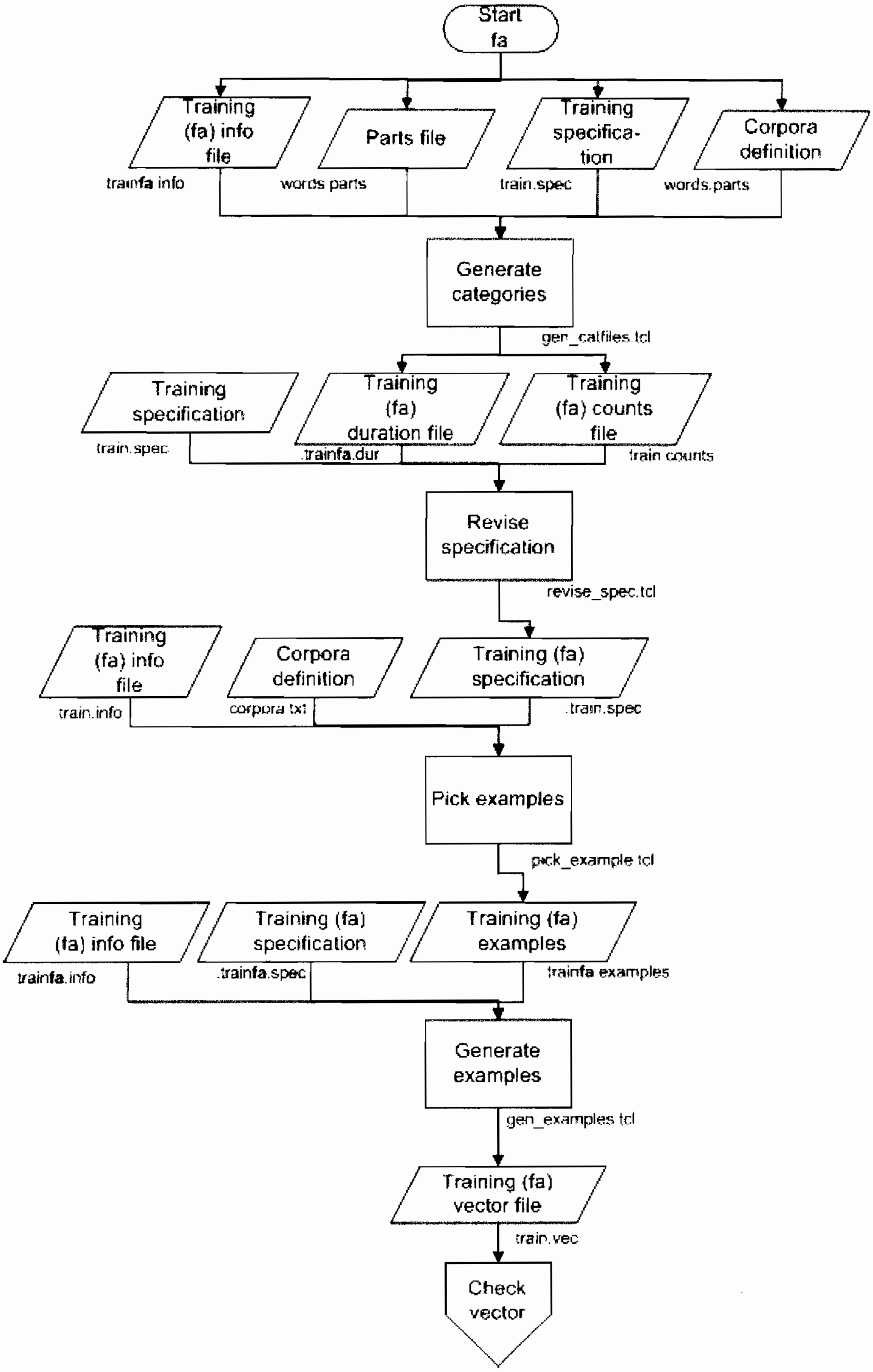


Figure 4.15. The force-alignment process.

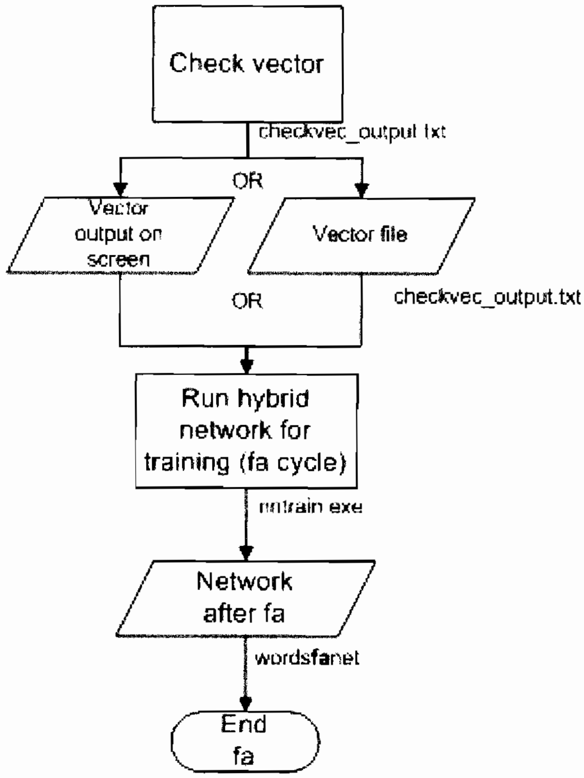


Figure 4.15 (cont.). The force-alignment process (continues).

4.4 Phoneme Refinement Strategy

The phoneme refinement strategy is emphasized to highlight its use in constructing the lexical model to achieve optimum WER. Phoneme refinement strategy involves an analysis of a word-to-word comparison at phoneme level. The comparison is performed on each target words of development dataset and the respective recognized words or hypotheses are aligned accordingly. For example see Table 4.4 (see Appendix F for a complete list). Hence, pronunciation variations at phoneme level, e.g. pronouncing the letter 'b' as the letter 'd' or 'p', is introduced into the lexicon by using the 'OR' operator similar to handling the miscues as pronunciation variations as discussed in section 4.2.1.

Table 4.4. Target words and recognized words (hypotheses) comparison.

Target word	**Hypotheses**
abang	*udang*
baca	*pada*
cendawan	*cindawan*
jangkitan	*janggitan*
jangan	*cendawan*

Phoneme refinement is performed prior to executing the data set. Phoneme refinement can be referred to as the process for allowing variations of pronunciation to be included in the pronunciation modeling. Thus, it is necessary to cope with the variability of dyslexic children's reading and spelling errors at phoneme level. This supports Noraini and Kamaruzaman (2008) that considers phoneme variations to increase recognition performance for standard BM. The phoneme variants considered in their study is based on place of articulation and manner or articulation of a phoneme.

It is observed that the optimum result of WER is obtained after phoneme refinement. The phoneme refinement rules, as presented in Table 4.5, are applied in the lexical model by allowing the phoneme variations of the letters to be included in the pronunciation modeling (see Figure 4.16 for illustration). The rule, adapted from Noraini and Kamaruzaman (2008), also includes the tendency to omit letters when reading by dyslexic children. The rule is based on place of articulation and manner of articulation (see Appendix G) with consideration to dyslexic children's tendency to remove a letter from a word (omission).

Table 4.5. The letters and their pronunciation variations allowed in the lexical model to enhance accuracy.

Character	Pronunciation variations
b	`p` OR `d` OR `m` OR omitted
a	`U`
e	`A`
j	`c` (vice versa)
k	`g` OR omitted
n	`ng` (vice versa)

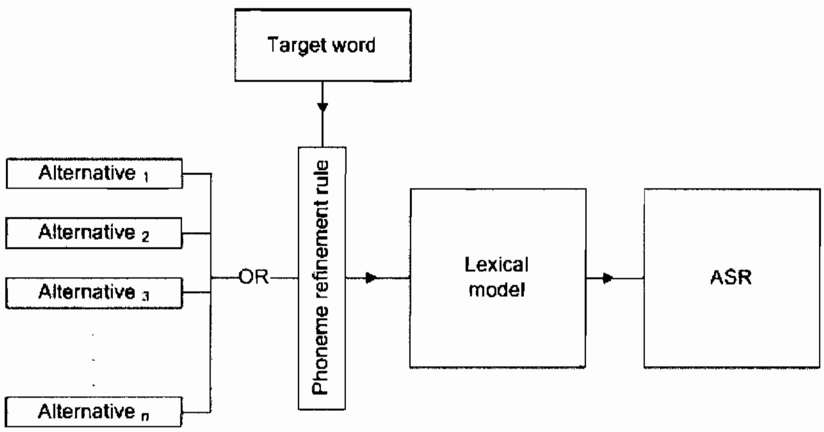


Figure 4.16. The illustration of phoneme refinement rule applied to modeling the lexicon.

4.5 Summary

With all the input descriptions and definition constructed, an ASR recognizer for dyslexic children's read speech is trained (see Chapter 5 for the results in details). Without a carefully modeled lexical model to cater for dyslexic children's reading isolated words in BM, an optimum accuracy could not be achieved. The lexicon is modeled based on data collection and analysis findings of the most frequent errors as presented in Chapter 3. The most frequent errors, not only support that phonological-core deficit is the major factor for reading difficulties for dyslexics but also act as the key towards increasing the recognition accuracy. Words with these errors are included in the lexical model and are modeled so that the recognizer allows for such pronunciation variations.

The proposed lexical model together with context-dependent modeling of pronunciations contributes to lower WER. Since this study deals with children's speech, which are known to be more challenging than that of adult's, context-dependent modeling is seen crucial to enable a more accurate classification of phonetic categories and thus, more accurate word recognition with lower WER.

Given that the lexicon consists of phonetically similar words due to their reading and spelling errors, the phonetic refinement by adapting pronunciation variations for dyslexic children has been incorporated into modeling the lexical model. Hence, this chapter has achieved the fifth objective of the study, which is to develop an ASR recognizer based on the ASR model proposed in Chapter 3. In the next chapter, the detail results are presented and the evaluation is discussed. The recognizer is evaluated in terms of the WER and in terms of its ability to detect miscues, measured in MDR.

CHAPTER 5

EVALUATION: WORD ERROR & MISCUE DETECTION

"How accurate is ASR and how can we tell? The general answer is "It depends" – in particular, on the function for which ASR is used, the purpose and criterion for evaluation, and the space in which evaluation is done."

(Mostow, 2006, p. 839-840)

5.0 Introduction

In Chapter 4, a recognizer is built to recognize dyslexic children reading selected vocabulary of isolated words in BM. Thus, WER is indeed an important measurement. The network that gives the best results of WER is regarded as the ASR recognizer as it gives the lowest error rate of 25%. Another criterion worth looking is the ability of the recognizer to detect miscues as this recognizer is meant for reading and thus should be able to detect miscues. Therefore, this chapter delineates the evaluation performed to evaluate the performance of the recognizer, which is developed based on the proposed ASR model. The evaluation is performed to evaluate the recognizer's performance on WER and MDR-FAR. The purpose of the evaluation is to answer whether or not the ASR recognizer is accurate enough given the scope-bounded vocabulary.

The performance of an ASR depends on many factors such as vocabulary size, amount of data for training, the type of words, and channel variability (J. –P. Hosom, personal communication, March 25, 2009). Additionally, evaluating a speech recognizer also depends on the purpose it is built for as emphasized by Mostow (2006). Therefore, the function of the recognizer is to be able to recognize reading miscues when dyslexic children read aloud the vocabulary. The criteria for evaluation, as mentioned, are WER and MDR-FAR. The space that the evaluation is performed is on *text space*, which means that the evaluation depends on the ability to recognize each word either as correct or misread. Other spaces include speech space and time-domain space, which refer to the ability to recognize sequence of words and the ability to recognize sequence of words with respect to time respectively (see Chapter 2, Section 2.6).

This chapter thus presents results of evaluation for testing the recognizer on a selected dataset for measuring its WER and MDR-FAR. Section 5.1 evaluates the performance of the recognizer in terms of WER. Section 5.2 elaborates on the methods to test the recognizer while Section 5.3 presents the results. Section 5.4 continues to present the analysis and discussion on the evaluation results on both metrics and Section 5.5 confers the summary.

5.1 Evaluation and Results: WER

As mentioned in Chapter 2, WER is a standard and widely accepted metric that is easy to use. It measures how accurately the recognizer recognizes spoken/read words. WER gives the error rate (%) deviation from the total percentage of 100% as:

```
WER = 100 (total %) - recognition rate (%)
```

In the study, the recognition rate is presented as the output of `select_best.tcl` command when executed to test the recognition accuracy on development and test datasets. Once the recognizer has been developed, the command `select_best.tcl` is executed on development and test datasets to obtain the best network that gives the best results on both sets. First, the command is executed on development dataset and the result is given in Table 5.1. For every iteration, the output is in the form of a network file named `wordsnet` or `wordsfanet` for force-alignment. Since training is performed for 30 iterations, the output of each of the iterations is automatically named as `wordsnet.1`, `wordsnet.2`, `wordsnet.3` and so on. The same goes to `wordsfanet`. Note that since this study involves recognition at word level, the insertion error percentage (*Ins%*) and the deletion error percentage (*Del%*) is always zero.

Table 5.1. The output of `select_best.tcl` on development dataset, extracted from the summary file showing word level accuracy.

Summary file: `words.dev.bmwords.files`					
Starting iteration: 30					
Itr	*#Words*	*Sub%*	*Ins%*	*Del%*	*WrdAcc%*
30	53	49.06	0.00	0.00	50.94
29	53	50.94	0.00	0.00	49.06
28	53	50.94	0.00	0.00	49.06
27	53	52.83	0.00	0.00	47.17
26	53	47.17	0.00	0.00	52.83
25	53	52.83	0.00	0.00	47.17
24	53	56.60	0.00	0.00	43.40
...	...	...	...	...	...
3	53	66.04	0.00	0.00	33.96
2	53	73.58	0.00	0.00	26.42
1	53	75.47	0.00	0.00	42.53
Best result:	52.83	with network `wordsnet.26`			
Evaluated:	30	networks			

After training, the best accuracy is 52.83% as highlighted in Table 5.1. Hence, force-alignment is required to obtain optimum result. The force-alignment is performed by automatically generating phonetic labels and training is thus repeated. Once the force-alignment training is completed, the command `select_best.tcl` is executed again on the development dataset and the result is as shown in Table 5.2. Referring to Table 5.2, the best accuracy is given by `wordsfanet.22` of 66.04%. Again, force-alignment is performed to find an optimum WER. The force-alignment is performed once more and the results are presented in Table 5.3.

Table 5.2. The output of `select_best.tcl` on development dataset after force-alignment, extracted from the summary file showing word level accuracy.

Summary file: `words.dev.bmwords.files`					
Starting iteration: 30					
Itr	*#Words*	*Sub%*	*Ins%*	*Del%*	*WrdAcc%*
30	53	39.62	0.00	0.00	60.38
29	53	43.40	0.00	0.00	56.60
28	53	36.85	0.00	0.00	64.15
...	...	...	...	...	...
22	53	33.96	0.00	0.00	66.04
...	...	...	...	...	...
3	53	41.51	0.00	0.00	58.49
2	53	39.62	0.00	0.00	60.38
1	53	58.49	0.00	0.00	41.51
Best result:	66.04	with network `wordsfanet.22`			
Evaluated:	30	networks			

Table 5.3. The output of `select_best.tcl` on development dataset after second force-alignment, extracted from the summary file showing word level accuracy.

Summary file: `words.dev.bmwords.files`					
Starting iteration: 30					
Itr	*#Words*	*Sub%*	*Ins%*	*Del%*	*WrdAcc%*
30	53	26.42	0.00	0.00	73.58
29	53	24.53	0.00	0.00	75.47
28	53	22.64	0.00	0.00	77.36
...	...	...	...	...	...
3	53	41.51	0.00	0.00	58.49
2	53	49.06	0.00	0.00	50.94
1	53	52.83	0.00	0.00	47.17
Best result:	77.36	with network `wordsfa2net.28`			
Evaluated:	30	networks			

After the **second** force-alignment the result gives a promising percentage of word accuracy of 77.36% and a lower WER of 22.64% with network `wordsfa2net.28`, as highlighted in Table 5.3. For a precaution, force-alignment is performed again to see whether or not better percentage can be achieved. However, the best percentage of the third cycle of force-alignment dropped to 64.15% (WER 35.85%) as illustrated in Table 5.4.

Table 5.4. The output of `select_best.tcl` on development dataset after third force-alignment, extracted from the summary file showing word level accuracy.

Summary file: `words.dev.bmwords.files`					
Starting iteration: 30					
Itr	*#Words*	*Sub%*	*Ins%*	*Del%*	*WrdAcc%*
30	53	35.85	0.00	0.00	64.15
29	53	43.40	0.00	0.00	56.60
28	53	45.28	0.00	0.00	54.72
...	...	...	...	...	...
3	53	39.62	0.00	0.00	60.38
2	53	43.40	0.00	0.00	56.60
1	53	60.38	0.00	0.00	39.62
Best result:	64.15	with network `wordsfa3net.30`			
Evaluated:	30	networks			

Table 5.5. The best network for the development dataset fed into cycles of force-alignment.

Network	**WER (%)**
`wordsnet.26`	47.17
`wordsfanet.22`	33.96
`wordsfa2net.28`	22.64
`wordsfa3net.30`	35.85
Optimum net:	`wordsfa2net.28` (22.64%)

To summarize, refer to Table 5.5. It can be concluded that, the WER is at its optimum after the second force-alignment with 22.64%. Therefore, the `wordsfa2net.28` is regarded as the best network (recognizer) for the development dataset. Thus, it is further evaluated on the test dataset and the result is depicted in Table 5.6.

Table 5.6. The final output evaluated on test dataset.

Summary file: `words.test.bmwords.files` Starting iteration: 30					
Itr	*#Words*	*Sub%*	*Ins%*	*Del%*	*WrdAcc%*
28	48	25.00	0.00	0.00	75.00

The evaluation on the test dataset produces the final result of the recognizer, which gives the WER of 25%. Again, `select_best.tcl` is executed on the test dataset but this time it only evaluates on `wordsfa2net.28` as it gives the best result on the development set. Thus, `wordsfa2net.28` is regarded as the best recognizer for recognizing dyslexic children's reading selected vocabulary of BM that falls under the most frequent error patterns. Now the recognizer (`wordsfa2net.28`) is fit for further evaluation to measure its MDR.

5.2 Evaluation: MDR

To evaluate the recognizer's ability in detecting miscues, the alignment technique is used. Its method and calculation are as presented in Chapter 3, Section 3.4.2. Basically, a total of 100 speech files of the target words in the trained lexicon are used as input to the recognizer. They are randomly selected from the speech files which are not included in the train, development, or test datasets with the intention that they contain about equal amount of correct and incorrect reading so that the recognizer's ability to detect miscues can be measured. So, a total of 52 files contain miscues and another 48 files are correct readings. Note that, the reason for the inclusion of correctly read words for the evaluation is to find out whether or not the recognizer generates wrong judgments on correctly read words, i.e. false alarm, which is measured using FAR.

Each of the speech files is input to the `asr_filter.tcl` command that uses the proposed recognizer and the output serves as the hypothesis for the alignment technique, which is adopted from Banerjee et al. (2003a) as described in Chapter 3. The hypothesis refers to the recognized word as output by the recognizer. The alignment technique is performed manually by comparing the target word, the hypothesis generated by the recognizer, and the transcript transcribed from the particular speech file. Given the hypothesis, the decision of whether or not a miscue is detected or undetected is based on the rules as follows:

- Miscue detected – a miscue is detected whenever the hypothesis deviates from its target word, i.e. contains error. However the miscue need not necessarily matches the transcript exactly. Consider the target word *abang*. The transcribed transcript gives *adang*, which is a miscue. However, the recognizer recognizes it as *ada* (hypothesis), which is also a miscue. Hence a miscue is detected even though *ada* did not match *adang* as transcribed in the transcript.
- Miscue undetected – a miscue is undetected whenever the hypothesis matches the target perfectly whereas the transcript contains an error. In other words, the miscue is undetected whenever an incorrect reading (based on transcript) is recognized as correct reading.
- False alarm – a false alarm is flagged whenever the hypothesis does not match the target word whereas the transcript gives a correct reading. This means that a false alarm happens when the recognizer misjudged a correct reading (based on transcript) as a miscue.

5.3 Results: MDR

To measure MDR, the transcripts, targets, and hypotheses are tabled accordingly for comparison, which is based on the aforementioned rules. A snippet of the table is presented in Table 5.7 that illustrates the hypotheses as compared to their corresponding transcripts and target words. A full alignment is presented in Appendix H.

Table 5.7. The alignment of transcripts, targets, and hypotheses.

Transcript	Target	Hypothesis	
adang	*abang*	*ada*	Miscue detected
abang	*abang*	*abang*	
abak	*abah*	*abah*	Miscue undetected
ubah	*abah*	*ubah*	Miscue detected
apa	*apa*	*apa*	
baja	*baca*	*bapa*	Miscue detected
ayah	*ayat*	*aya*	Miscue detected
bawang	*bawang*	*wan*	False alarm
belalah	*belalang*	*abah*	Miscue detected

Based on the alignment table, the number of miscues detected, miscues undetected, and false alarm is calculated. A total of 42 miscues were detected, where 10 of them were undetected and a number of 8 correct readings were falsely flagged by the recognizer as miscues. Again, as defined in Chapter 2: 1) MDR is given by the number of detected miscues divided by the total number of miscues made; and 2) FAR is given by the number of correct readings falsely recognized as incorrect by the recognizer divided by the total number of correct readings. Hence, the MDR and FAR of the recognizer are given by:

$$\text{Miscue detection rate, MDR} = \frac{42}{52} \times 100\% = 80.77\%$$

$$\text{False alarm rate, FAR} = \frac{8}{48} \times 100\% = 16.67\%$$

The MDR obtained is promising and outperforms some of the speech recognizers related to children's reading (see Chapter 2, Section 2.6). The FAR too is considered to be within the estimated figure (around 15% – 30% as shown in Table 2.5) as reviewed in Chapter 2. However, it is acknowledged that the tradeoffs between MDR and FAR always exist where higher MDR yields higher FAR (Banerjee et al., 2003a; Lee et al., 2004; Tam et al., 2003; Duchateau et al., 2007; Li et al., 2008; Liu et al., 2008). Similar to the MDR obtained, Lee et al. demonstrated that when the MDR reached an optimum of 80%, the false alarm increased up to 36.4%. Given the 'nature' of the vocabulary involved in the recognizer that is highly phonetically similar, achieving 80.77% MDR and 16.67% FAR is indeed a significant performance.

5.4 Analysis and Discussion

The accuracy of 80.77% MDR could not be achieved should the recognizer fail to settle at the optimum WER. Besides, achieving the MDR could not be achieved if the lexical model is not modeled as such and emphasized in the proposed ASR model. Modeling pronunciation variations as alternative pronunciations and phoneme refinement did contribute to achieving the WER and MDR-FAR. Hence, they contribute to the performance of the recognizer. This section is meant for analysis of the effects of the proposed lexical model to the recognizer's accuracy (WER and MDR).

5.4.1 The Effects of Phoneme Refinement on WER, MDR, and FAR.

Without modeling the errors as pronunciation variants with phoneme refinement, higher WER is obtained. Table 5.8 and Table 5.9 illustrate the results of the recognizer **without** alternative pronunciations and phoneme refinement for development dataset and test dataset respectively. This recognizer is regarded as a counter model to the proposed lexical model. The counter model is constructed of the same lexical representation but **without** phoneme refinement strategy. The shaded iteration 10 (Table 5.8) represents the best recognition results obtained of 52.54% giving WER of 47.46% on development dataset. Since the network in iteration 10 (`wordsfanet.10`) presents the best accuracy on the development dataset, the network is used to evaluate the recognizer's performance on test dataset and the result is presented in Table 5.9. Note that, even though the network on iteration 1 produces the same percentage as in iteration 10, the network on iteration 10 is chosen because it is calculated and discovered first before iteration 1 (as the cycle iterates from iteration 30 to 1, i.e. in a decreasing manner).

Table 5.8. The output of `select_best.tcl` on development dataset.

Summary file: `words.dev.bmwords.files`					
Starting iteration: 30					
Itr	*#Words*	*Sub%*	*Ins%*	*Del%*	*WrdAcc%*
30	53	52.54	0.00	0.00	47.46
29	53	54.24	0.00	0.00	45.76
...	...	...	...	...	...
10	53	47.46	0.00	0.00	52.54
...	...	...	...	...	...
2	53	55.93	0.00	0.00	44.07
1	53	47.46	0.00	0.00	52.54
Best result:	52.54	with network `wordsfanet.10`			
Evaluated:	30	networks			

Table 5.9. The output of `select_best.tcl` on test dataset.

Summary file: `words.test.bmwords.files` Starting iteration: 30					
Itr	*#Words*	*Sub%*	*Ins%*	*Del%*	*WrdAcc%*
10	48	38.18	0.00	0.00	61.82

As emphasized, the lower the WER is the better. To lower the WER, higher recognition accuracy has to be achieved. Therefore, as suggested by Noraini and Kamaruzaman (2008), phoneme refinement is able correct the recognition errors and thus the phoneme refinement is applied to the proposed lexical model in order to increase recognition accuracy (see Chapter 4, Section 4.4). Once applied, its performance is compared.

In order to evaluate the performance of the proposed lexical model **with** phoneme refinement, the counter model is used. The comparison is as illustrated in Table 5.10 (includes results on development dataset for recognition rate and WER) and Figure 5.1. The performance is compared in terms of WER, MDR, and FAR. The comparison is made to demonstrate the significant performance of the proposed model with regards to its counter model (refer to Appendix I for the miscue detection alignment of the recognizer built on the counter model). Note that, the alignment technique is also applied to calculate MDR and FAR for the proposed model after the first and third force-alignment to measure the performances (see Appendix J).

Table 5.10. The performance of the proposed lexical model and its counter model measured in WER, MDR, and FAR.

	WER%		MDR%	FAR%
	Dev. set	**Test set**		
Model **without** phoneme refinement and alternative pronunciation	47.46	38.18	82.69	35.42
Model **with** phoneme refinement and alternative pronunciation	22.64	25.00	80.77	16.67

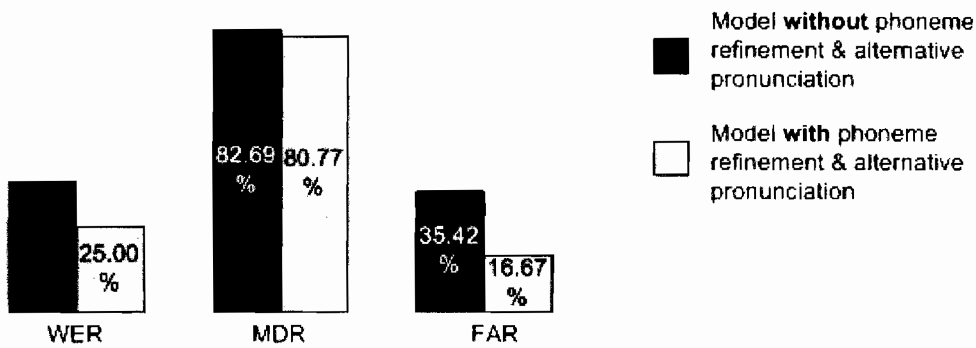


Figure 5.1. The final performance comparison between different lexical models of the same data, measured in WER and MDR-FAR.

Referring to Figure 5.1, it can be concluded that the overall performance in terms of WER, MDR, and FAR for the proposed model is significantly better, given the same training data, the same amount of training data, the same development and test datasets, and the same miscue detection data. The WER is measured using the test dataset on the final network as discussed in Chapter 4. The inclusion of the proposed lexical model into the ASR model manages to lower the WER by 13.18%. Although the MDR for the proposed model is relatively lower (only 1.92%) than that of the counter model, it still is the best performance taking into account the FAR. The counter model performs worse with more than double increase in FAR as opposed to the proposed model.

With considerably high MDR the recognizer has demonstrates its reliability in detecting miscues. In addition, the proposed lexical model of the ASR model for dyslexic children manages to lower the FAR by 18.75%. The performance indicates the trustworthiness of the recognizer to perform its tasks – to accurately recognize read words and detect miscues with a considerable rate for false alarm detection. As asserted, in any automated reading tutor, miscue detection is a key feature so that correction procedure could be triggered for useful feedback to users. Hence, MDR-FAR can also be regarded as a way to measure the reliability of a recognizer meant for recognizing read speech and detecting reading miscues.

5.4.2 The Overall Performance

Overall, the ASR recognizer has a promising potential to recognize dyslexic children's reading selected vocabulary in BM. Since the WER of 25% is optimum, the recognizer has demonstrate its ability to recognize dyslexic children's reading with a lot of phonological errors, mostly vowel and consonant substitutions, errors related to nasals ('m' and 'n'), and consonant omissions.

In addition, with the MDR as high as 80.77% the recognizer has also demonstrate its ability to reliably detect miscues whenever the children read a word incorrectly. Table 5.11 illustrates the performance of the proposed recognizer, which is in line with other recognizers built to detect reading miscues. It is emphasized that the illustration of performances of various recognizers is not meant for performance comparison but rather to demonstrate current similar recognizer's MDR and FAR. A fair comparison is seen implausible as each of the recognizers used different amount of speech data, different test data, different vocabulary, and even different language to be recognized, as discussed in Chapter 2, Section 2.6. Since the proposed recognizer is the only recognizer dealing with highly phonetically vocabulary of dyslexic children, its performance can be regarded as optimum.

Table 5.11. The performance of the proposed recognizer among the recognizers built for similar purposes.

	User type	**Language**	**Evaluation Metrics**	
			MDR (%)	**FAR (%)**
Banerjee et al. (2003a)	Normal children	English	42.53	2.90
Tam et al. (2003)	Normal children	English	58.64	2.92
Hagen (2006)	Normal children	English	73.00	3.00
Li, Deng, Ju, & Acero (2008)	Normal children	English	76.90	15.80
Li et al. (2007)	Normal children	English	76.93	15.15
Lee et al. (2004)	Normal children	English	80.00	36.40
The proposed recognizer	Dyslexic children	Bahasa Melayu	80.77	16.67
Duchateau et al. (2007)	Normal children	Dutch	83.10	8.40
Liu et al. (2008)	Normal children	Mandarin	88.80	11.50

As emphasized, the MDR is essentially important for a recognizer meant for reading recognition as it could guide/trigger suitable correction feedback. If accurate recognition is demonstrated by a recognizer, correct output can be displayed and suitable feedback can be triggered. With that, this recognizer has the potential to be further applied into automatic reading tutor system because its ability to produce correct recognition and most importantly, its ability to detect reading errors.

5.4.3 The Shortcoming

Even though the recognizer demonstrates an optimum performance in terms of WER and MDR, the false alarm needs to be minimized. It has been observed that the speech files that contribute to false alarm recognitions contain rather strong background noise signal that clearly affect the recognition task. For speech files that contain little or no background noise, the recognizer manages to produce either correct recognition or detected the miscues as demonstrated by the WER and the MDR respectively. As such, one of the shortcomings demonstrated by the recognizer is that it works better in a controlled environment where background noise is reduced to its minimum, as most recognizers are.

Another shortcoming is that it is not capable of handling dialects and regards dialects as errors. For example, when a child read *hampar* as 'hampaq' the recognizer shall identify it as an error. However, often when reading aloud, standard BM is used that does not include dialects – as defined in the pronunciation models of the words. Still, 'hampaq' is regarded as a consonant substitution error, substituting the letter 'r' with 'q' and thus, it is recognized as a miscue. Although most of the participants did not read *hampar* as hampaq or *umur* as 'umoq', still it is an open problem that needs to be handled intelligently so that these dialect-influenced pronunciations can be treated accordingly.

5.5 Summary

It can be concluded that given the WER of 25%, the recognizer, which has been developed with dyslexic children's most frequent errors presented in its lexical model, give a significant performance. It manages to detect miscues with the MDR of 80.77% and the FAR of 16.67%, having misrecognized only 8 correct readings as incorrect out of a total of 48 correct readings. The significant performance shows that the recognizer is reliable even when dealing with phonetically similar words.

Without the modeling of phonetically similar words (the errors) as alternative pronunciations together with phoneme refinement, such accuracy could not be achieved – be it the WER or the MDR and FAR. After all, these three metrics measures how accurate an ASR is, whether it fits the task of recognizing words with phonetically similar errors or not and whether it could generate reliable MDR and FAR. Thus, it is proven that the lexical model with pronunciation variations and phoneme refinement acts as an important component in an ASR for dyslexic children's reading BM controlled vocabulary. That is why the lexical model with pronunciation variations and phoneme refinement is highlighted in the proposed model of ASR in Chapter 3.

With that, this chapter has answered the last question – can the model be evaluated and what are the criteria to take into consideration when evaluated. In conclusion, the criteria for evaluation are the ASR accuracy measured in WER and the accuracy of detecting reading miscues measured in MDR and FAR, as demonstrated. Answering the question means that the final objective has been fulfilled that is to evaluate the proposed ASR model in terms of WER, MDR, and FAR. The final chapter concludes the research and discusses its limitations and future directions.

CHAPTER 6

CONCLUSION & FUTURE DIRECTIONS

6.0 Introduction

The previous five chapters have structurally presented the research from outlining the objectives and the theoretical background of the study to the ASR recognizer's development and evaluation. This chapter concludes the thesis by presenting the objectives that have been successfully achieved. Basically, this chapter re-established the objectives, the methods, and the deliverables of the study. The research is recapitulated highlighting how it relates to the phonological deficit theory and connects it to an enhanced ASR model. In addition, future works are also outlined for future research directions in ASR and dyslexic children in the same or similar domain.

To emphasize the overall research performed to complete the thesis, this chapter summarizes them in the following sections: Section 6.1 recaps the research by illustrating the phonological-core deficit theory and how it relates to the enhanced ASR model for dyslexic children. Section 6.2 lists all six objectives that have been achieved by describing, in brief, each objective's methods and deliverables; Section 6.3 emphasizes the contribution of the thesis to the field of ASR as well as education/special education language-related field, specifically dyslexia and BM; Section 6.4 directs some future research directions; and Section 6.5 summarizes the final chapter of the thesis.

6.1 Research Recapitulation

The phonological-core deficit theory is regarded as the major factor to contributing dyslexia. The discussion of this theory is presented in Chapter 2. According to this theory, what causes reading mistakes are the broken link(s) of the cognitive processes involved in reading. Due to the broken links that connect one cognitive process to another, one or more reading mistakes are fabricated. These mistakes are of phonological-based errors and thus, these errors are often of words that are phonologically similar. For example, *ayat* is incorrectly read as *ayah* or *aya* and *selesa* is incorrectly read as *Selasa* as demonstrated in this study. Therefore the error patterns support that phonological-core deficit theory is indeed the key contributing factor to this difficulties, producing phonological reading errors such as vowel and consonant substitutions, nasals, and consonant omission.

The phonological reading errors are modeled into the reading model (the word recognition model) so that the most frequent reading error patterns are incorporated into the cognitive processes. It is important that the reading process and the errors they produce are presented visually for a clearer picture of the effects dyslexia has upon reading. The need to model the errors is also contributed by less knowledge of the errors that dyslexia causes when reading in BM, as mentioned. Furthermore, it is crucial to have this knowledge in order to pursue a fine-tuned dyslexic-based ASR model so that recognition accuracy is not compromised. Thus, as reviewed in Chapter 2, many suggested that the mispronunciations are included into the lexical model of the ASR.

Unfortunately, the phonetically similar words or the reading errors could damage the accuracy of a speech recognizer because ASR is poor at recognizing fine phonetic details for example, trying to differentiate between the sounds of bh and ph (J.-P. Hosom, personal communication, May 14, 2008). This situation will definitely have a negative effect on the recognition accuracy. Furthermore, the research involving children's speech is known to be more difficult.

To handle the challenges, the ASR model proposed carefully modeled the mispronunciations into its active lexicon as pronunciation variations with the inclusion of phonetic refinement. The lexical model is then trained using the hybrid HMM/ANN method for developing a recognizer. The recognizer is further evaluated using the test dataset, to obtain an optimum result. Thus with the promising WER of 25%, it is proven that the developed recognizer based on the ASR model proposed, which contains the specifically modeled lexical model, leverages the challenge that phonetically similar words imposed on ASR. Figure 6.1 summarizes the research.

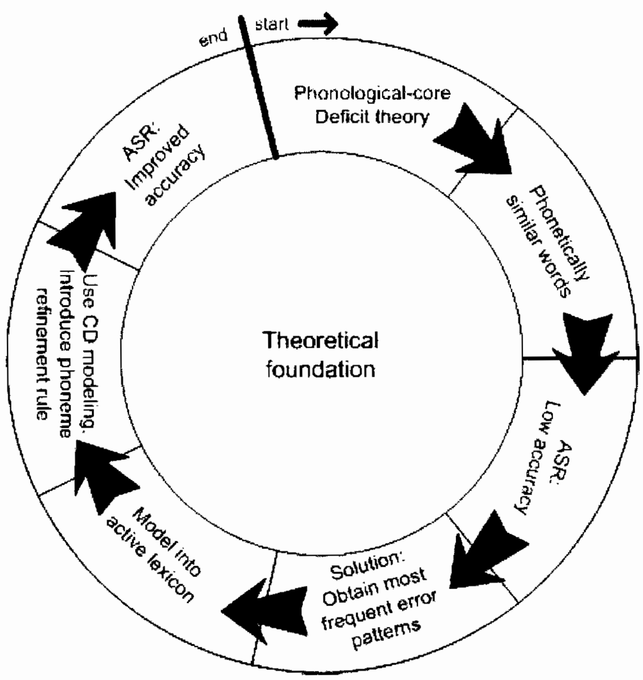


Figure 6.1. The theoretical foundation constitutes to dyslexia and phonetically similar vocabulary and how it benefits the ASR.

This study enhances the phonological-core deficit theory in two aspects: 1) it supports that the phonological-core deficit theory is indeed the major cause of dyslexia among BM readers; 2) it suggests that the theory can also be applied to ASR for the significant information it provides to enhancing recognition accuracy for dyslexic children's reading in BM. In particular, the information regarding the reading errors surfaced from the analysis of the study provides insights as how to model the errors such that ASR accuracy is not compromised and can be increased.

These enhancements are supported by two important factors: 1) that the findings of data analysis present similar phonological-based errors as of English (Sawyer, et al., 1999); and 2) that the theory could be used to establish a direct connection between dyslexia and ASR in terms of providing a dyslexic-tuned lexical model that handles phonetically similar vocabulary and thus increasing the recognition accuracy for an improved performance.

6.2 The Objectives Achieved

The main objective defined earlier in Chapter 1 Section 1.4 is to propose an ASR model that is designed based on dyslexic children's reading model in BM. Therefore, six specific objectives have been defined to achieve the main objective and each objective that has been fulfilled is summarized in the following subsections.

The objectives answer to the six questions mentioned earlier in Chapter 1. The questions are: 1) is there any specific vocabulary for dyslexic children to learn to read?; 2) what are the reading and spelling patterns for dyslexic children reading in BM?; 3) can a reading model be developed to model dyslexic children reading in BM?; 4) can an

ASR model be modeled based on the previous reading model?; 5) how can the model be designed and developed?; 6) how can the model be evaluated and what are the criteria to take into consideration when evaluated? Table 6.1 summarizes the objectives, the results and deliverables, and the contributions made.

Table 6.1. The results and deliverables and their contributions outlined for the specific objectives achieved.

Objective	**Results/Deliverables**	**Contributions**
To collect vocabulary in BM	114 single, isolated words	A reading model that models the most frequent reading errors of dyslexic children's reading in BM
To recognize reading mistakes and discover dyslexic children's reading and spelling patterns	16 BM-based reading mistakes by dyslexic children	
To model dyslexic children's reading in BM	Dyslexic children's reading model for BM	
To model an ASR for the reading model	ASR model for dyslexic children (BM)	An enhanced hybrid HMM/ANN ASR model that emphasized on the lexical model (incorporation of the most frequent errors and phonetic refinement into the lexical model)
To develop an ASR recognizer based on the ASR model	ASR recognizer	An ASR recognizer as a working model that demonstrates its ability to recognize readings of selected words in BM with acceptable rates.
To evaluate the proposed ASR model for recognition accuracy (WER and MDR, FAR)	Evaluation results – accuracy (%) in WER and MDR (taking into account the FAR)	

6.2.1 To collect vocabulary in BM

The vocabulary composes the lexicon for a speech recognizer. What really matters in this case is to collect suitable vocabulary that is within the dyslexic children's reading level and contained in the syllabus. As mentioned in Chapter 1, the reading level of concern is at word recognition level, which is the early readers' level that involves single word reading. Therefore single, isolated words in BM were selected based on suggestions made by the discussion with special education teachers.

The discussion lead to a fruitful list of single words within the primary school syllabus of level one (*tahap satu*). A total of 114 words were selected for data collection to record and obtain dyslexic children's readings for the purpose of collecting the speech signals and analyzing the errors. The words are of variety of syllable patterns in BM including prefixes and postfixes forms of consonant and vowel combinations, e.g. 'me-', 'pen-', and '-an'.

The words serve as the vocabulary of which the children were asked to read aloud into a head mounted microphone for recording purposes. The speech signals were collected as evidence of reading and spelling errors or miscues that surfaced from reading the single words. What make this a challenging process was the response from the children participated in the study whose difficulties in reading hinder them from paying considerable concentration on the instructions and task at hand. However, the readings managed to be successfully recorded for analysis and thus fulfilled the first objective of the study.

The first objective answers the first question regarding suitable vocabulary for dyslexic children. Some suggestions of suitable vocabulary by other works are presented in Chapter 2. To answer the question, Chapter 3 presents and discusses the methods to collect the vocabulary for reading error analysis.

6.2.2 To recognize reading and spelling patterns

It is important to highlight that the reading mistakes do provide richer information as to how to deal with dyslexic children's miscues when developing a recognizer specifically targeted for their reading. The reason is that the incorporation of the miscues into the lexical model can help increase its accuracy significantly as demonstrated in Chapter 5.

The analysis performed discovered the reading and spelling error patterns as thoroughly discussed in Chapter 3. Mainly, the analysis involved studying the pronunciations of dyslexic children and deducing the type of errors, which are adapted from Sawyer et al. (1999). The error patterns discovered share similar characteristics where vowel substitution is found to be the most frequent error followed by consonant omission and nasals (errors with substitutions and deletions of 'm' and 'n'), and consonant substitutions. Although BM and English are two different languages, it seems that the most frequent error reflects that vowels are indeed the most confusing letters to the children in order to produce correct phones or sounds.

With the analysis that yielded significant most frequent error patterns as mentioned, the second objective is achieved and thus answers to the second question. It is an important contribution that the errors, which surfaced from BM readings, are recognized in order to probe deeper understanding and motivate future works on this field. A thorough discussion on dyslexia and reading errors is reviewed in Chapter 2. An in depth discussion on recognizing and discovering the error patterns is presented in

Chapter 3. Achieving this objective means that the error patterns need to be used as the key towards achieving the following objectives.

6.2.3 To model dyslexic children's reading in BM

To exclusively model a reading model for dyslexic children dedicated for BM requires visualizing its broken links of cognitive processes involved. So to visualize them, the four most frequent errors – vowel substitution, consonant omission, nasals, and consonant substitution – are introduced into the existing word recognition model to emphasize the effect(s) dyslexia has on reading BM.

The incorporation of the error types into the word recognition model, which is adapted from Ellis (1993), also presents a cause-effect relationship between the cognitive processes and the errors that they produce once the connection(s) is/are impaired. The justification to adapt Ellis's word recognition model is as explained in Chapter 2. The link from visual analysis module to visual input lexicon module to speech output lexicon module might be impaired and thus produce the errors when reading. Such broken links could cause failure to decipher the written text presented and affect other cognitive processes thus producing these errors. Visual analysis, visual input lexicon, and speech output lexicon modules are the processes involved in reading as explained in Chapter 2.

For this reason, the incorporation of the most frequent errors is considered as an enhancement of the current model for a more dyslexic-tuned reading model. With that, the third objective of the study has been achieved and therefore answers the third question of the reading model for dyslexic children. The detailed methods are presented in Chapter 3.

6.2.4 To model an ASR for the reading model

The adaptation of the errors into the reading model provides information as to how to model ASR that is fine-tuned towards dyslexic children's reading. It thus invoked the question of how ASR can deal with the correct pronunciations and mispronunciations. The answer leads to the lexical model where adaptation of the errors into active lexicon can increase accuracy when they are modeled as pronunciation variations of the target words.

The lexical model or the pronunciation models of the vocabulary, which includes the miscues, is further enhanced in an existing ASR model. The model adapts that of HMM/ANN model for reasons as discussed earlier in Chapter 2. The justification is that the hybrid model tends to perform better as compared to the state-of-the-art HMM as reviewed. After all, accuracy is the utmost important aspect when ASR is of concern. Thus, with the incorporation of the most frequent errors into the lexical model of the hybrid ASR model, the fourth objective is successfully achieved.

Achieving the fourth objective answers the fourth question of the research. The detailed methods are presented in Chapter 3 together with the proposed model (see Figure 3.6).

6.2.5 To develop an ASR recognizer based on the ASR model

In conjunction with the ASR model, an ASR recognizer of hybrid network is developed to evaluate the proposed model. The development of the ASR recognizer is performed using CSLU Toolkit. The toolkit ease the development process in which it allows to focus on building the lexical component that is tuned towards modeling the errors so that recognition accuracy can be increased. To better enhance the performance, phoneme refinement is also introduced as a filter to further refine certain letters such as 'b' and 'j' as described in Chapter 4, Section 4.4.

To develop, the speech files (reading recordings) collected together with their corresponding transcription and phoneme label files are automatically divided accordingly so that the training dataset contains more files, about 3/5 of the data. Therefore the development and test dataset contain only 1/5 of speech files each. A detail flow of the training process commenced is presented in Figure 4.14 and 4.15 of Chapter 4. Basically it illustrates the input and output of a process needed to be performed during training.

The development is a success when the developed recognizer is evaluated using the test dataset where it generates an optimum result a considerably low WER of 25% with 75% recognition rate. Therefore the fifth objective has been fulfilled and answers the fifth question of developing the model into a working recognizer. A thorough discussion on the development process is presented in Chapter 4.

6.2.6 To evaluate in WER and MDR

As mentioned, the recognizer developed managed to achieve significant accuracy of 25% WER even though it deals with phonetically similar words. This is due to the carefully modeled lexical model that plays an important role towards achieving such performance. Consequently, the lexical model cannot be modeled as such without the important information that both the reading model and ASR model have to offer.

Due to the fact that the ASR model and the recognizer is fine-tuned towards dyslexic children's reading, it therefore must be able to detect reading miscues. To measure the performance, MDR is also calculated to measure the accuracy of miscue detection. When miscue detection is of concern, it directly relates to measuring false alarm. Thus, FAR is also measured to support MDR as the two metrics are normally being applied together to obtain the performance of a reading-oriented ASR. Hence, the recognizer scored 80.77% MDR and acceptable 16.67% FAR that demonstrates that the recognizer is a success. For that, the final objective has been achieved, answering to the sixth question of whether or not the model can be evaluated. The full results and analysis is presented in Chapter 5.

6.3 Upshot – The reading model and the ASR model

The overall study suggests that for an ASR to recognize phonetically similar words rooted from the phonological-core deficit theory of dyslexia, the lexical model has to be emphasized and constructed accordingly to occupy the phonetically similar errors and refine certain letters. Thus, the significant contributions are: 1) the reading model based on dyslexic children's reading in BM; 2) the ASR model for dyslexic children's reading; and 3) the recognizer based on the ASR model proposed.

Firstly, the need to first construct the reading model of BM in the context of dyslexic children raised from the apparent less knowledge of dyslexia and BM and its effects on the language. It is therefore an important contribution as it enriches the knowledge of dyslexia and its effects on the cognitive processes of reading, as well as its consequences to BM. The reading model also provides vital information for ASR in terms of what to emphasize and how to do so with the intention that accuracy is not compromised. The dyslexic reading model focuses on the effects of deteriorated links of cognitive processes that suggests that the most frequent errors as an outcome of the processes (see Figure 3.5 of Chapter 3). The outcome is held as an important element to be included when proposing an ASR model for dyslexic children to ensure optimum accuracy. Thus, the proposed ASR model incorporates these errors with emphasize on its lexical component as presented in Figure 3.6 of Chapter 3. Clearly, the reading model is indeed a contribution to knowledge of dyslexia and its relation to BM.

Secondly, it also contributes to the knowledge of modeling pronunciations for lexical model in ASR so that better and increased accuracy can be achieved as demonstrated in Chapter 5. With the knowledge of the errors in BM, the ASR model is proposed that is fine-tuned towards the readings of dyslexic children. As mentioned, the proposed ASR model emphasizes on its lexical component targeted at achieving high accuracy. Thus, the ASR model is certainly a contribution to knowledge in the field of ASR where it provides fundamental presentation of the lexical model to be incorporated in a BM-based reading-oriented ASR for dyslexic children.

Finally, the developed recognizer is regarded as a novel effort to implement the ASR model and measure its performance in terms of accuracy and its ability to detect reading miscues. Thus, the recognizer serves as means to evaluate the ASR model and acts as a baseline for a BM-based reading-oriented recognizer.

The aforementioned contributions offer benefits to various beneficiaries such as dyslexic children, teachers, parents, researchers, software developers, and the MoE. The reading model enriches the knowledge to understand more on how dyslexia affects the cognitive processes in reading. The model benefits the teachers in that it provides the information on the most frequent reading error patterns in BM. Teachers as well as researchers can focus on methods to reduce and eliminate such errors when teaching the dyslexic children to read. Thus, the dyslexic children could benefit from a more BM-oriented teaching method in order to overcome their difficulties. Since special education is in focus, which has gained serious attention from the government in particular the MoE, teachers or reading facilitators, as well as parents whose children are diagnosed as dyslexics could gain the benefits too.

The ASR model and its developed recognizer could benefit the researchers in related fields to further enhance the model into a working BM-based automatic reading tutor. Researchers and software developers could further improve the ASR recognizer to cater for more vocabulary and sentence structure. The dyslexic children could enhance their reading skill and improve reading performance by using such reading tutor that 'listens' to their reading and give corrective feedbacks. Thus, teachers and parents could use the reading tutor and provide support for their students and children to learn to read in BM.

6.4 Future works

Further effort could be performed to enhance this study in terms of enhancing the mechanism to collect data (corpus collection), vocabulary, continuous recognition as well as coping with dialect-influenced pronunciations.

But first of all, it is worth to look at how to make data collection an effective process. During data collection process, the children were asked to read aloud the selected words into a microphone and the readings were recorded for further analysis. The process of reading recording was performed manually – present a participant with a word, ask the participant to read it aloud, record the reading simultaneously, and save the recorded reading into the computer. It is discovered that manual data collection of such is very time consuming and sometimes frustrating especially when dealing with a dyslexic child struggling to read even a simple word. Hence, automatic data collection can ease the process.

The automatic data collection could make the reading recording faster and more fun to the children and indirectly increase their motivation to perform the task at hand – reading. The fun part can be instilled to the children when they can see an interesting cartoon-like interface that pops up or display each word that they are going to read. It is important to keep in mind that these children have little motivation when reading and thus find reading a boring process. Using computer-based application specifically designed and built to record their reading and at the same time displaying interesting interface could be a great help to researcher.

The recognizer only models controlled vocabulary that falls under the four most frequent errors. The vocabulary is also limited to only words of which the errors occur only once in a particular word. Therefore, adding more words into the vocabulary could increase the ASR robustness for it can recognize more words and perhaps large vocabulary ASR can be considered. The work presented models only isolated words in BM since the target users/participants are of primary school dyslexic children constitutes to the first level of reading, which is the word level reading. Adding more words to the vocabulary could also consider targeting students with sentence level reading skill.

When it involves sentence level reading, continuous recognition is also another option that can be considered. In fact, continuous recognition suits passage reading for dyslexic children reading at sentence level, which caters for the second level of reading (*tahap dua* in primary schools). The work can be extended to cater for continuous ASR by carefully modeling its language model that maps the grammar of sentences in BM. When reading isolated words, grammar is not the major focus. It is indeed important when continuous recognition is of concern as it provides syntactic knowledge to ASR.

Besides the aforementioned future directions, efforts to handle dialect-influenced pronunciations could also be taken. The dialect-influenced pronunciations, as explained in Chapter 5, are currently recognize as incorrect/miscues. Works that can be considered to eliminate this limitation include considering dialect-specific modeling in the lexical model so that the dialect could be accepted by the recognizer. An example work on dialect-specific modeling is as presented by Livescu and Glass (2000).

6.5 Summary

Even with the limitations as mentioned, the research is a breakthrough in ASR dedicated for BM and dyslexic children, linking the phonological-deficit theory, reading in BM and ASR. The research proposes a reading model that is fine-tuned for dyslexic children reading in BM where it highlights the four most frequent error patterns observed when the children read aloud selected vocabulary. The error patterns serve as the key to model a dyslexic-tuned lexical model. The lexical model carefully models the reading errors of dyslexic children so that recognition accuracy is not compromised.

The ASR model proposed emphasizes on the lexical model as an important component when it comes to recognition of dyslexic children's reading. A filter is proposed into the model to cater for phoneme refinement rule as an enhancement of the lexical model to boost recognition accuracy. Prior to the research, there is no evidence of knowledge of the dyslexic children's most frequent errors when reading in BM. Also, there is no evidence of an ASR model that is fine-tuned for reading recognition in BM for dyslexics. Thus, the research contributes to the enrichment of knowledge in the aforementioned fields.

To conclude, it is asserted that the ASR model serves a fundamental model for building reading-oriented ASR for BM. The model is constructed with emphasized on modeling its lexical model as a key element in the ASR process so that the miscues can be manipulated to increase accuracy. Thus, the thesis presented a novel approach to present and model the reading errors in an ASR model designated for isolated word recognition, in particular the BM-based automatic reading tutor for dyslexic children.

References

Aist, G., Chan, P., Huang, X. D., Jiang, L., Kennedy, R., Latimer, D., et al. (1998). How effective is unsupervised data collection for children's speech recognition? *Proceedings of the International Conference on Speech and Language Processing (ICSLP98)*, Sydney, Australia, 3171-3174.

Association for Computing Machinery. (1998). *The ACM Computing Classification System 1998*. Retrieved Feb 14, 2008, from http://www.acm.org/class/1998/ccs98.html.

Audiblox. (2000). *Phonological awareness and phonemic awareness*. Retrieved Mar 30, 2007, from http://www.audiblox2000.com/learning_disabilities/phonological.htm.

Bain, K., Basson, S. H., & Wald, M. (2002). Speech recognition in university classrooms: Liberated Learning Project. *Proceedings of the 5th International ACM Conference on Assistive Technologies*, Edinburgh, Scotland, 192 -196.

Banerjee, S., Beck, J., & Mostow, J. (2003a). Evaluating the effect of predicting oral reading miscues. *Proceedings of the 8th International European Conference on Speech Communication and Technology (Eurospeech-03)*, Geneva, Switzerland, 3165-3168.

Banerjee, S., Mostow, J., Beck, J., & Tam, W. (2003b). Improving language models by learning from speech recognition errors in a reading tutor that listens. *Second International Conference on Applied Artificial Intelligence*, Fort Panhala, Kolhapur, India.

Baumer, B. H. (1998). *How to teach your dyslexic child to read: A proven methods for parents*. New York: Kensington Publishing Corp.

Bourlard, H., & Morgan, N. (1998). Hybrid HMM/ANN systems for speech recognition: Overview and new research directions. In C.L. Giles & M. Gori (Eds.), *Adaptive Processing of Sequences and Data Structures, Lecture Notes in Artificial Intelligence (1387),* (pp. 389-417). USA: Springer Verlag.

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read: A causal connection. *Nature, 301,* 419-421.

Castles, A. (2006). The dual-route model and the developmental dyslexia. *London Review of Education, 4*(1), 49-61.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel distributed processing approaches. *Psychological Review, 100*(4), 589-608.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual-route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*(1), 204-256.

Conn, N., & McTear, M. (2000). Speech technology: A solution for people with disabilities. *IEEE Seminar on Speech and Language Processing for Disabled and Elderly People,7,* London, UK, 1-6.

Cosi, P. (2000). Hybrid HMM-NN architectures for connected digit recognition. *IEEE Transactions on ASSP, 5,* 85-90.

Das, S., Nix, D., & Picheny, M. (1998). Improvements in children's speech recognition performance. *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing 1998,* Seattle, WA, USA, *1,* 433-436.

Duchateau, J., Wigham, M., Demuynck, K., & Van hamme, H. (2006). A flexible recognizer architecture in a reading tutor for children. *Proceedings of ITRW on Speech Recognition and Intrinsic Variation*, Toulouse, France, 59-64.

Dupont, S., Ris, C., Couvreur, L., & Boite, J.-M. (2005). A study of implicit and explicit modeling of coarticulation and pronunciation variation. *Proceedings of Interspeech 2005*, Lisbon, Portugal.

Dwyer, B. (2000). *The uses of computer technology in the remediation of children with specific learning difficulties (Dyslexia).* Retrieved Mar 28, 2007, from http://webpages.dcu.ie/~farrenm/spec.pdf

Dyslexia Association Singapore, DAS. (2003). *DAS hypothesis on the cause of dyslexia.* Retrieved August 14, 2007, from http://www.das.org.sg/news/dashyp.htm

D'Arcy, S. M., Wong, L. P., & Russell, M. J. (2004). Recognition of read and spontaneous children's speech using two new corpora. *Proceedings of International Conference on Spoken Language Processing (ICSLP'04)*, Jeju Island, Korea.

Ellis, A. W. (1993). *Reading, writing and dyslexia: A cognitive analysis* (2nd ed.). UK: Lawrence Erlbaum Associates Ltd.

Fairweather, P., Nix, D., Oblinger, D., Adams, B., & Carla, L. (n.d.). *Overcoming technical barriers to a speech-enabled children's reading tutor.* Retrieved Mar

13, 2007, from http://www.research.ibm.com/AppliedLearningSciWeb/Fairweather/techbar.pdf

Fausett, L. V. (1994). *Fundamentals of neural networks: Architectures, algorithms, and applications*. New Jersey: Prentice-Hall.

Fawcett, A. J., & Nicolson, R. I. (2002). Children with dyslexia are slow to articulate a single speech gesture. *Dyslexia, 8*, 189-203.

Franco, H., Cohen, M., Moran, N., Rumelheart, D., & Abrash, V. (1994). Context-dependent connectionist probability estimation in a hybrid hidden Markov model-neural net speech recognition. *Computer Speech and Language, 8,* 211-222.

Frost, J. (2001). Phonemic awareness, spontaneous writing, and reading and spelling development from a preventive perspective. *Reading and Writing: An Interdisciplinary Journal, 14,* 487 - 513.

Goswami, U. (2002). Phonology, reading development and dyslexia: A cross-linguistic perspective. *Annals of Dyslexia, 52,* 1-23.

Hagen, A. (2006). *Advances in children's speech recognition with application to interactive literacy tutor*. Unpublished doctoral dissertation, University of Colorado, Colorado.

Hagen, A., Pellom, B., & Cole, R. (2003). Children's speech recognition with application to interactive books and tutors. *Proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding 2003*, St. Thomas, USA, 186-191.

Hagen, A., Pellom, B., Van Vuuren, S., & Cole, R. (2004). Advances in children's speech recognition within an interactive literacy tutor. *Proceedings of HLT-NAACL*, Boston Massachusetts, USA.

Hain, T. (2002). Implicit pronunciation modelling in ASR. *ISCA Tutorial and Research Workshop on Pronunciation Modeling and Lexicon Adaptation for Spoken Language Technology (PMLA)*, Colorado, USA, 129-134.

Hieronymus, J. L. (1993). *ASCII phonetic symbols for world's languages: Worldbet* (Tech. Rep.). NJ, USA: AT&T Bell Labs.

Higgins, E. L. (2004). Speech recognition-based and automaticity programs to help students with severe reading and spelling problems. *Annals of Dyslexia, 54,* 365 - 392.

Higgins, E. L., & Raskind, M. H. (2000). Speaking to read: The effects of continuous vs. discrete speech recognition systems on the reading and spelling of children with learning disabilities. *Journal of Special Education Technology, 15,* 19 - 30.

Hosom, J.P., Cole, R.A, & Cosi, P. (1999). Improvements in neural-network training and search techniques for continuous digit recognition, *Australian Journal of Intelligent Information Processing Systems (AJIIPS), 5*(4), 277-284.

Hosom, J. P., de Villiers, J., Cole, R., Fanty, M., Schalkwyk, J., Yan, Y. et al. (2006). Training Hidden Markov Model/Aritificial Neural Network (HMM/ANN) hybrids for automatic speech recognition (ASR) [Computer software and manual]. Retrieved July 7, 2007, from http://www.cslu.cse.ogi.edu/tutordemos/

Husniza, H., & Zulikha, J. (2010). Improving ASR performance using context-dependent phoneme models. *Journal of Systems and Information Technology (JSIT), 12*(1).

Husniza, H. & Zulikha, J. (2009a). Pronunciation variations and context-dependent model to improve ASR performance for dyslexic children's read speech. *Proceedings of International Conference on Computing and Informatics (ICOCI 2009)*, Kuala Lumpur, Malaysia.

Husniza, H. & Zulikha, J. (2009b). Dyslexic Children's Reading Pattern as Input for ASR: Data, Analysis, and Pronunciation Model. *Journal of Information and Communication Technology (JICT), 8,* 1-13.

Indirawati, Z. & Mardian, S. O. (2006). *Fonetik dan Fonologi Bahasa Melayu.* Kuala Lumpur, Malaysia: PTS Professional Publishing.

International Dyslexia Association. (2006). *What is Dyslexia?* Retrieved Mar 30, 2007, from http://www.interdys.org/servlet/compose?section_id=5&page_id=95

Jabatan Pendidikan Khas. (1999). *Buku Panduan Pelaksanaan Program Pemulihan Khas: Masalah Penguasaan 3M.* Kuala Lumpur, Malaysia: Jabatan Pendidikan Khas, Kementerian Pelajaran Malaysia.

Jelinek, F. (2001). *Statistical methods for speech recognition.* Cambridge, Massachusetts: The MIT Press.

Kasselimis, D., Margarity, M., & Vlachos, F. (2006). The relation of dyslexia to cerebellar function and articulation speed. *Annals of General Psychiatry, 5*(1), 160.

Kementerian Pelajaran Malaysia. (2006). *Pelan Induk Pembangunan Pendidikan (PIPP) – Edisi Pelancaran PIPP 2006*. Kuala Lumpur, Malaysia: Kementerian Pelajaran Malaysia (KPM).

Lee, K. A., Hagen, A., Romanyshyn, N., Martin, S., & Pellom, B. (2004). Analysis and detection of reading miscues for interactive literacy tutors. *Proceedings of the 20th International Conference on Computational Linguistics (Coling) 2004*, Geneva, Switzerland.

Lerner, J. (1997). *Learning disabilities: Theories, diagnosis, and teaching strategies* (7th ed.). Boston: Houghton Mifflin Company.

Li, X., Ju, Y., Deng, L., & Acero, A. (2007). Efficient and robust language modeling in an automatic children's reading tutor systems. *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'07)*, Honolulu, Hawaii, IV-193-IV-196.

Li, X., Deng, L., Ju, Y., & Acero, A. (2008). Automatic reading tutor on hand held devices. *Proceedings of the Interspeech 2008*, Brisbane, Australia, 1733-1736.

Liu, C., Pan, F., Ge, F., Dong, B., Zhao, Q., & Yan, Y. (2008). Application of LVCSR to the detection of Chinese Mandarin reading miscues, *Proceedings of International Conference on Natural Computation, 5*, 447-451.

Liu, M., Xu, B., Huang, T., Deng, Y., & Li, C. (2000). Mandarin accent adaptation based on context-independent/context-dependent pronunciation modeling. *Proceedings of the IEEE International Conference on Acoustic, Speech, and Signal Processing, 2*, II1025 - II1028.

Livescu, K. & Glass, J. (2000). Lexical modeling of non-native speech for automatic speech recognition. *Proceedings of the IEEE International Conference on Acoustic, Speech, and Signal Processing,* Istanbul, Turkey, *3*, 1683-1686.

Lundberg, I. (1995). The computer as a tool of remediation in the education of students with reading disabilities: A theory-based approach. *Learning Disability Quarterly, 18*(2), 88-99.

Lundberg, I. & Olofsson, A. (1993). Can computer speech support reading comprehension? *Computers in Human Behavior, 9,* 283-293.

Manis, F. R., Custodio, R., & Szeszulski, P. A. (1993). Development of phonological and orthographic skill: A 2-year longitudinal study of dyslexic children. *Journal of Experimental Child Psychology, 56,* 64-86.

Markowitz, J. A. (1996). *Using speech recognition.* New Jersey: Prentice Hall PTR.

Md Sah, S., Dzulkifli, Y., & Sheikh Hussein, S. S. (2001). Neural network speaker dependent isolated Malay speech recognition system: Handcrafted vs. genetic algorithm. *Proceedings of International Symposium on Signal Processing and Its Applications (ISSPA)*, Kuala Lumpur, 731-734.

Mioduser, D., Tur-Kaspa, H., & Leitner, I. (2000). The learning-value of computer based instruction of early reading skills. *Journal of Computer Assisted Learning*, *16*, 54-63.

Mostow, J. (2006). Is ASR accurate enough for automated reading tutors, and how can we tell? *Proceedings of the 9th International Conference on Spoken Language Processing (Interspeech 2006 — ICSLP).* Pittsburgh, 837-840.

Mostow, J., Roth, S., Hauptmann, A. G., & Kane, M. (1994). A prototype reading coach that listens. *Proceedings of the 12th National Conference on Artificial Intelligence (AAAI-94)*, Seattle, WA, 785-792.

Mostow, J., Beck, J., Winter, S. V., Wang, S., & Tobin, B. (2002). Predicting oral reading miscues. *Proceedings of the 7th International Conference on Spoken Language Processing* (ICSLP-02), Denver, CO, 1221-1224.

Mostow, J., & Beck, J. (2003). When the rubber meets the road: Lessons from the in-school adventures of an Rutomated Reading Tutor that listens. *Proceedings of the Conceptualizing Scale-Up: Multidisciplinary Perspectives*, Park Hyatt Hotel, Washington, D.C.

National Dissemination Center for Children with Disability. (2004). *Learning Disabilities*. Retrieved Mar 7, 2007, from http://www.nichcy.org/pubs/factshe/fs7.pdf

Nix, D., Fairweather, P., & Adams, B. (1998). Speech recognition, children, and reading. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, Los Angeles, USA, 245-246.

Noraini, S., & Kamaruzaman, J. (2008). Acoustic pronunciation variations modeling for standard Malay speech recognition. *Journal of Computer and Information Science, 1*(4), 112-120.

Nor Hasbiah, U. (2007). *Perisian Kursus Multimedia dalam Literasi Matematik (D-Matematika) untuk Pelajar Disleksia*. Unpublished doctoral dissertation, Universiti Kebangsaan Malaysia, Bangi.

Olofsson, A. (1992). Synthetic speech and computer aided reading for reading disabled children. *Reading and Writing: An Interdisciplinary Journal, 4,* 165-178.

Olson, R. K. & Wise, B. W. (1992). Reading on the computer with orthographic and speech feedback: An overview of the Colorado remediation project. *Reading and Writing: An Interdisciplinary Journal, 4,* 107-144.

Rabiner, L. & Juang, B-H. (1993). *Fundamentals of Speech Recognition.* New Jersey: PTR Prentice-Hall, Inc.

Rainger, P. (2003). *A dyslexic perspective on e-content accessibility.* Retrieved Mar 9, 2007, from http://www.techdis.ac.uk/resources/files/dyslexia.doc

Ranaivo-Malacon, B. (2005). Malay lexical analysis through corpus-based approach. *Persidangan Antarabangsa Leksikologi dan Leksikografi Melayu (PAMA 2005),* Kuala Lumpur, Malaysia.

Raskind, M. H., & Higgins, E. L. (1999). Speaking to read: The effects of speech recogition technology on the reading and spelling performance of children with learning disabilities. *Annals of Dyslexia, 49,* 251 - 281.

Raskind, M. H., & Shaw, T. (1999). Assistive technology for individuals with learning disabilities. *Proceedings of Technology and Persons with Disabilities Conference 1999.* Retrieved Mar 9, 2007, from http://www.csun.edu/cod/conf/1999/ proceedings/session1002.htm

Reid, G. (2003). *Dyslexia: A practitioner's handbook* (3rd ed.).West Sussex: John Wiley & Sons Ltd.

Renals, S., Morgan, N., Bourlard, H., Cohen, M., & Franco, H. (1994). Connectionist probability estimators in HMM speech recognition. *IEEE Transaction on Speech Audio Processing, 2,* 1, 161-174.

Rigoll, G., & Willett, D. (1998). A NN/HMM hybrid for continuous speech recognition with a discriminant nonlinear feature extraction. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP),* Seattle, USA, 9-12.

Russell, M., Brown, C., Skilling, A., Series, R., Wallace, J., Bonham, B., et al. (1996). Applications of automatic speech recognition to speech and language development in young children. *Proceedings of the 4th International Conference on Spoken Language ICSLP'96,* Philadelphia, USA, *1,* 176-179.

Sawyer, D. J., Wade, S., & Kim, J. K. (1999). Spelling errors as a window on variations in phonological deficits among students with dyslexia. *Annals of Dyslexia, 49,* 137-159.

Seidenberg, M. S., & McClelland, J. M. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*(4), 523-568.

Shaywitz, S. E. (1996, November). Dyslexia. *Scientific American,* 98-104.

Sheikh Hussain, S. S., Ahmad, Z., Zulkarnain, Y., Rahman, S., & Lim, S. C. (2000). Implementation of speaker identification systems by means of personel computer. *Proceedings of TENCON 2000,* Kuala Lumpur, Malaysia, *1,* 43-48.

Shobaki, K., Hosom, J. P., & Cole, R. A. (2000). The OGI KIDS' Speech corpus and recognizers. *Proceedings of the 6th International Conferenceon Spoken Language Processing (ICSLP),* Beijing, China.

Smith, L., Pham, T., Bower, J., Pilgrim, B., Schwarz, P., & Stevens, D. (2002). Toward voice applications for children. *International Journal of Speech Technology, 5,* 321-329.

Snowling, M. J. (2000). *Dyslexia* (2nd ed.). UK: Blackwell Publishers.

Stanovich, K. E., & Seigel, L. S. (1994). Phenotypic performance profile of children with reading disability: A regression-based test of the phonological-core variable-difference model. *Journal of Educational Psychology, 86,* 24-53.

Syed Abdul Rahman, A-H., Salina, A. S., Aini, H., Khairul, A. I., & Mirvaziri, H. (2007). Decision fusion for isolated malay digit recogition using dynamic time warping (DTW) and Hidden Markov Model (HMM). *Proceedings of the 5th Student Conference on Research and Development (SCOReD 2007).*

Sutton, S., Cole, R. A., de Villiers, J., Schalkwyk, J., Vermeulen, P., Macon, M. et al. (1998). Universal speech tools: The CSLU toolkit. *Proceedings of the 5th International Conference on Spoken Language Processing,* Sydney, Australia, *7,* 3221-3224.

Tam, Y.-C., Mostow, J., Beck, J., & Banerjee, S. (2003). Training a confidence measure for a reading tutor that listens. *Proceedings of the 8th European Conference on Speech Communication and Technology*, Geneva, Switzerland, 3161-3164.

Ting, H. N., Jasmy, Y., Sheikh Hussain, S. S., & Cheah, E. L. (2001). Malay syllable recognition based on Multilayer Perceptron and Dynamic Time Warping. *Proceedings of International Symposium on Signal Processing and Its Applications (ISSPA)*, Kuala Lumpur, Malaysia, 743-744.

Ting, H. N., Jasmy, Y. & Sheikh Hussein, S. S. (2001). Speaker-independent Malay syllable recognition using singular and modular neural networks. *Jurnal Teknologi, 35,* 65-76.

Trentin, E., & Gori, M. (2001). A survey of hybrid ANN/HMM models for automatic speech recognition. *Neurocomputing, 37,* 91-126.

Trentin, E., & Gori, M. (2003). Robust combination of neural network and hidden Markov models for speech recongition. *IEEE Transactions on Neural Network, 14,* 6, 1519-1531.

Tsakalidis, S., Prasad, R., & Natarajan, P. (2009). Context-dependent pronunciation modeling for Iraqi ASR. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2009*, 4457-4460.

University of Washington. (2002). *DO-IT - Working together: Computers and people with learning disabilities.* Retrieved Mar 18, 2007, from http://www.washington.edu/ doit/Brochures/ PDF/atpwld.pdf

Vellutino, F. R., Scanlon, D. M., & Spearing, D. (1995). Semantic and phonological coding in poor and normal readers. *Journal of Experimental Child Psychology, 59,* 76-123.

Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, A., Chen, R. S. et al. (1996). Cognitive profiles of difficult-to-remediate and readily remediate poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology, 88*(4), 601-638.

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific Reading Disability (dyslexia): What we have learned in the past four decades?, *Journal of Child Psychology and Psychiatry, 45*(1), 2-40.

Wald, M. (2004). Using automatic speech recognition to enhance education for all students: Turning a vision into reality. *Proceedings of the 34th ASEE/IEEE Frontiers in Education Conference,* Indianapolis, Indiana, USA, 22-25.

Wald, M. (2005). SpeechText: Enhancing learning and teaching by using automatic speech recognition to create accessible synchronised multimedia. *Proceedings of ED-MEDIA 2005 World Conference on Educational Multimedia, Hypermedia, and Telecommunications*, Montreal, 4765-4769.

Wald, M. (2006). An exploration of the potential of automatic speech recognition to assist and enabled receptive communication. *ALT – J: Research in Learning, 14*(1), 9 - 20.

Williams, S. M., Nix, D., & Fairweather, P. (2000). Using speech recognition technology to enhance literacy instruction for emerging readers. In B. Fishman & O'Connor-Divelbiss (Eds.). *Fourth International Conference of the Learning Sciences*, 115-120, Mahwah, NJ: Erlbaum.

Wolf, M. (1999). What time may tell: Towards a new conceptualization of developmental dyslexia. *Annals of Dyslexia, 49*, 3-28.

Wolf, M., Bowers, P. G., & Biddle, K. (2000). Naming speed processes, timing, and reading: A conceptual review. *Journal of Learning Disabilities, 33*(4), 387-407.

Yan, Y., Fanty, M., & Cole, R. (1997). Speech recognition using neural networks with forward-backward probability generated targets. *IEEE International Conference* *on Acoustics, Speech, and Signal Processing, ICASSP-97,* Munich, Germany, *4,* 3241-3244.

Zaharin, Y. (2000). Computational linguistic in Malaysia. *Proceedings of the 38th Annual Meeting on Association of Computational Linguistics,* Hong Kong, 1-2.

Ziegler, J. (2006). Do differences in brain activation challenge the universal theories of dyslexia? *Brain and Language, 98*, 341-343.

Zulikha, J., & Abdullah, E. (2007). *A Malay speech recognition prototype for telephony-related vocabulary* (Tech. Rep.), UUM Sintok, Kedah, Malaysia: Universiti Utara Malaysia.

Appendix A

Consonant–Vowel (CV) pairs in syllable patterns of BM words

Source: *Buku Panduan Pelaksanaan Program Pemulihan Khas (Masalah Penguasaan 3M)*

Syllable patterns in words	Example
V+CV	*ibu*
CV+CV	*meja*
CV+CV with digraph	*wangi*
CV+CV with diphthong	*gurau*
CVC	*bas*
V+CVC	*enam*
V+CVCC	*orang*
CV+CVC or CVC+CV	*lapan*
CV+CVCC	*burung*
CV+CVC or CVC+CV	*pantai*
CVC+CVC	*sampah*
CVCC+CVCC	*pinggang*
Dual vowels	*buas*
CV+CV+CV	*kerusi*
CV+CV+CCV	*telinga*
CV+CV+CVV	*serunai*
CV+CV+CCVV	*perangai*
CVC+CV+CVC	*cendawan*
CV+CV+CVCC	*binatang*

CV+CVC+CVCC	*melambung*
CVC+CV+CVCC	*pembohong*
CVCC+CV+CVC	*tengkorak*
CV+CVCC+CVC	*menangkap*

Appendix B

The Readings of Dyslexic Children According to Word Syllable Patterns

Syllable pattern: V+CV

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
aku	DC1	aku	aku	aku	aku	aku	aku	aku
	DC2	aku	aku	aku	aku	aku	aku	aku
	DC3	aku	aku	aku	aku	aku	aku	aku
	DC4	aku	aku	aku	aku	aku	aku	aku
	DC5	aku	aku	aku	ku	aku	aku	aku
	DC6	aku	aku	-	-	-	-	-
	DC7	aku	aku	aku	aki	aku	apa	aku
	DC8	aku	aku	aku	aku	aku	aku	aku
	DC9	aka	aka	-	-	-	-	-
	DC10	aku	aku	aku	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
itu	DC1	itu	itu	itu	itu	itu	itu	itu
	DC2	itu	itu	itu	itu	itu	itu	itu
	DC3	itu	itu	itu	itu	itu	itu	itu
	DC4	itu	itu	itu	itu	itu	itu	itu
	DC5	itu	itu	itu	itu	itu	itu	itu
	DC6	itu	itu	-	-	-	-	-
	DC7	itu	itu	itu	itu	itu	ini	itu
	DC8	itu	itu	itu	itu	itu	itu	itu
	DC9	itu	itu	-	-	-	-	-
	DC10	itu	itu	itu	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
apa	DC1	apa	apa	apa	apa	apa	apa	apa
	DC2	apa	apa	apa	apa	apa	apa	apa
	DC3	apa	apa	apa	apa	apa	apa	ape
	DC4	ape	ape	ape	ape	ape	ape	ape
	DC5	apa	apa	apa	apa	apa	apa	apa
	DC6	apa	apa	-	-	-	-	-
	DC7	apa	apa	apa	apa	apa	apa	apa
	DC8	apa	apa	apa	apa	apa	apa	apa
	DC9	apa	apa	-	-	-	-	-
	DC10	apa	apa	apa	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
ibu	DC1	ibu	ibu	ibu	ibu	ibu	ibu	ibu
	DC2	ibu	ibu	ibu	ibu	ibu	ibu	ibu
	DC3	ibu	ibu	ibu	ibu	ibu	ibu	ibu
	DC4	ibu	ibu	ibu	ibu	ibu	ibu	ibu
	DC5	ibu	ibu	ibu	ibu	ibu	ibu	ibu
	DC6	ibu	ibu	-	-	-	-	-
	DC7	ibu	ibu	ibu	ibu	ibu	ibu	ibu
	DC8	ibu	ibu	ibu	ibu	ibu	ibu	ibu
	DC9	ibu	ibu	-	-	-	-	-
	DC10	ibu	ibu	ibu	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
ada	DC1	ada	ada	ada	ada	ada	ada	ada
	DC2	ada	ada	ada	ada	ada	ada	ada
	DC3	bapak	aba	ada	ada	ada	ada	ada
	DC4	ada	ade	ade	ade	ade	ade	ade
	DC5	ada	ada	ada	ada	ada	ada	ada
	DC6	ada	ada	-	-	-	-	-
	DC7	ada	ada	ada	ada	ada	ada	ada
	DC8	ada	ada	ada	ada	ada	pada	ada
	DC9	ada	ada	-	-	-	-	-
	DC10	ada	ada	ada	-	-	-	-

Syllable pattern: CV+CV

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
beli	DC1	beli	beli	beli	beli	beli	beli	beli
	DC2	eli	beri	beli	beli	beli	beli	beli
	DC3	beli	beli	beli	beli	beli	beli	beli
	DC4	beêdi	beli	beli	beli	beli	beli	beli
	DC5	beli	beli	beli	beli	beli	beli	beli
	DC6	beri	beli	-	-	-	-	-
	DC7	beli	beli	beli	beli	beli	beli	beli
	DC8	beli	beli	beli	beli	beli	beli	beli
	DC9	beli	peli	-	-	-	-	-
	DC10	beli	beli	beli	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
pada	DC1	pada	pada	pada	bapak	bapak	pada	pada
	DC2	pada	pada	pada	pada	pada	pada	pada
	DC3	pa	bapa	pada	bapa	bapa	bapa	bapa
	DC4	ada	pada	bapa	pade	pade	pada	pada
	DC5	pada	pada	pada	pada	pada	pada	pada
	DC6	pada	pada	-	-	-	-	-
	DC7	bata	pada	pada	ada	ada	papak	bapak
	DC8	pada	pada	pada	pada	pada	pada	bapa
	DC9	pada	pada	-	-	-	-	-
	DC10	pada	pada	pada	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
baca	DC1	baca	baca	baca	baca	baca	baca	baca
	DC2	baca	baca	baca	baca	baca	baca	baca
	DC3	baca	baca	baca	beca	baca	baca	baca
	DC4	bace	baja	baca	baca	bace	baca	baca
	DC5	baca	baca	baca	baca	baca	baca	baca
	DC6	baca	baca	-	-	-	-	-
	DC7	baci	bica	baca	paca	baca	bapak	baca
	DC8	baca	baca	baca	baca	baca	baca	baca
	DC9	baka	baca	-	-	-	-	-
	DC10	baca	baca	baca	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
bapa	DC1	apa	bapak	bapa	bapak	bapak	bapak	bapajk
	DC2	dapa	dapa	bapa	dapa	bapa	bapa	bapa
	DC3	apa	bapa	bapak	bapa	apa	bapa	bapa
	DC4	bapak	bapak	bapak	bapak	bapak	bapak	babak
	DC5	pada	pada	patah	pada	bada	bapa	pada
	DC6	apa	bapa	-	-	-	-	-
	DC7	bapa	bapa	bapa	bapak	bapak	bapak	papa
	DC8	pada	bapa	bapa	bapa	bapa	bapa	bapa
	DC9	ada	ada	-	-	-	-	-
	DC10	dapa	dapa	dapa	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
suka	DC1	suka	duka	suka	suka	suka	suka	suka
	DC2	suka	suka	suka	suka	suka	suka	suka
	DC3	suka	suka	suka	suka	suka	suka	suka
	DC4	suka	suka	suka	suka	suke	suka	suka
	DC5	suka	suka	suka	suka	suka	suka	suka
	DC6	suka	suka	-	-	-	-	-
	DC7	suka	suka	suka	suka	suka	suka	suka
	DC8	suka	suka	suka	suka	suka	suka	suka
	DC9	suka	sukar	-	-	-	-	-
	DC10	suka	suka	suka	-	-	-	-

Syllable pattern: CV+CV with digraph

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
punya	DC1	punya	punya	punya	unya	punya	punya	punya
	DC2	puya	pun_ya	pun_ya	punya	pun_ya	pun_ya	pun_ya
	DC3	-	punya	punya	punya	punya	punya	punya
	DC4	punya	punya	humnya	punya	punya	punya	punya
	DC5	pannya	puya	punya	pun_ya	puya	punnya	punnya
	DC6	pun_ya	bun_ya	-	-	-	-	-
	DC7	punya	punya	punya	punya	punnya	punya	punya
	DC8	punang	punya	punya	penya	punya	paya	punya
	DC9	patanya	punnya	-	-	-	-	-
	DC10	punnya	punya	puye'	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
sunyi	DC1	sunyi	sunyi	sunyi	sunyi	sunyi	sunyi	sunyi
	DC2	sunyi	sun_wi	sun_ya	sun_yi	sunyia	sunyi	sun_yi
	DC3	sembunyi	bunyi	sembunyi	sembunyi	sembunyi	sembunyi	sembunyi
	DC4	suni	suni	sunyi	sunyi	sunyi	sunyi	sunyi
	DC5	suni	suyi	suni	sunyi	suni	suni	suni
	DC6	sun_yi	sun_yi	-	-	-	-	-
	DC7	sunyi	sunnya	suni	sunyi	sunyi	sunyi	sunyi
	DC8	suya	suna	sunyi	sunya	sunyi	sunyi	sunyi
	DC9	saya	sunyi	-	-	-	-	-
	DC10	suni	suyi	sunyi	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
tanya	DC1	tanya	tanya	tanya	tanya	tanya	tanya	tanya
	DC2	tan_ya	tan_ya	tan_ya	tan_ya	tan_ya	tan_ya	tan_ya
	DC3	tanya	tanya	tanya	tanya	tanya	tanya	tanya
	DC4	tanya	tanya	tanya	tanya	tanya	ta	tanya
	DC5	tan_ya	tan_ya	tanya	tanya	tanya	tanya	tanya
	DC6	tinya	tan_ya	-	-	-	-	-
	DC7	tanya	tanya	tannya	tunnya	tannya	tunnya	tanya
	DC8	tanya	tenang	tanya	tanya	tanya	tanya	tanya
	DC9	siya	tan	-	-	-	-	-
	DC10	taya	tanya	taya	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
nyata	DC1	nyati	anatan	nyata	nyata	enyata	anyata	nyata
	DC2	ya	nyata	nyata	nyata	nyata	nyata	nyata
	DC3	-	nukmat	hayati	hayati	-	nyata	nyata
	DC4	nyata	nyata	nyata	menata	nyata	nyata	nyata
	DC5	nyata	nyata	nyatah	nyata	nyata	nyata	nyata
	DC6	nata	-	-	-	-	-	-
	DC7	nyata	nyata	nyata	nyata	nyata	nyata	tanya
	DC8	nyata	nangta	nyata	nyata	nyata	nyata	nyata
	DC9	naya	anya	-	-	-	-	-
	DC10	ayatan	ayat	ayat	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
bunga	DC1	bun_ga	bunga	bunga	bunga	bunga	bunga	bunga
	DC2	bun_ga	bun_ga	bun_gan	punka	punkan	bunga	bunga
	DC3	bunga	bunga	bunga	bunga	bunga	bunga	bunga
	DC4	bunge	bunga	unah	buna	buna	buna	bunga
	DC5	bunga	buna	bunga	bungga	bunga	bunga	bunga
	DC6	tanga	bunja	-	-	-	-	-
	DC7	buji	bunja	benja	banda	buga	buna	buga
	DC8	bena	bujang	bunga	bunga	bunga	bunga	boja
	DC9	banyak	abang	-	-	-	-	-
	DC10	buga	bugan	buga	-	-	-	-

Syllable pattern: CV+CV with diphthong

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
gurau	DC1	gurau	gurau	suruh	guru	guru	juru	juru
	DC2	-	guru	juru	gurau	gurong	guru	gura
	DC3	-	guru	gurun	gurau	gurun	gurun	gurun
	DC4	gurau	gurau	gurau	gurau	gurau	gurau	boru
	DC5	guar	gurau	gurau	gurau	gurau	gurau	gurau
	DC6	-	-	-	-	-	-	-
	DC7	jurah	juru	pura	jura	guru	juru	burang
	DC8	gurang	curang	gurau	cura	gurau	guru	juru
	DC9	surau	guli	-	-	-	-	-
	DC10	gurang	guran	guru	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
serai	DC1	silam	serai	serai	serai	seri	seri	serai
	DC2	seran	serang	serang	serang	serang	serang	serai
	DC3	seri	siri	serai	surai	serai	serai	cerai
	DC4	seri	seri	se'ri	saru	seri	serang	seri
	DC5	cerah	serar	sewai	serar	serai	serai	serahai
	DC6	seri	reni	-	-	-	-	-
	DC7	surah	sira	suyi	seri	sura	seri	suri
	DC8	sureng	serai	serai	sera	serar	seri	sera
	DC9	sukar	ais	-	-	-	-	-
	DC10	serai	serai	serai	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
kelui	DC1	keleli	kurai	kuli	kuri	keri	kekuli	kuli
	DC2	-	kelur	kelur	kelu	kelu	kelur	keliu
	DC3	puli	kuli	bikul	keldai	kulai	kuli	kuli
	DC4	kuli	kelayu	kelu	kuli	kuli	kelayu	kelu
	DC5	kelari	kelawah	kelauoi	kelawor	kelau	kelauwar	kelawar
	DC6	kayut	kuli	-	-	-	-	-
	DC7	-	kiku	kela	kela	kelu	keli	kelu
	DC8	kelui	keluar	keluar	keru	keluar	keluar	keluli
	DC9	lalu	kili	-	-	-	-	-
	DC10	kelui	kelu	kelu	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
wahai	DC1	wahai	wahai	wahid	wahai	wahai	wahai	wahai
	DC2	waha	waha	waha	wahai	wahai	wahai	wahai
	DC3	wari	wadi	walhal	waihan	wani	wahani	wahai
	DC4	wahi	wahi	wahi	wah	wahi	wahi	wah
	DC5	wahai	wawah	wahai	wahai	awai	wahai	wahai
	DC6	wahi	wahi	-	-	-	-	-
	DC7	wahi	wahi	wahi	wahi	wahi	wahi	wahi
	DC8	wahi	wahi	wahi	wahai	wasir	wahai	waha
	DC9	wahai	awan	-	-	-	-	-
	DC10	wahi	wahai	wahai	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
ceria	DC1	ceria	ceria	ceria	ceria	ceria	ceria	ceria
	DC2	cira	ceria	ceria	ceria	ceria	ceria	ceria
	DC3	curia	ceria	cari	ceria	ceria	ceria	ceria
	DC4	ceria	ceria	ceria	ceria	ceria	ceria	ceria
	DC5	ceria	ceria	ceria	ceria	ceria	ceria	ceria
	DC6	cera	cerin	-	-	-	-	-
	DC7	cerah	kiri	cera	kerna	kerna	ceri	ringgi
	DC8	caria	ceria	ceda	ceria	ceria	ceria	cera
	DC9	cari	cili	-	-	-	-	-
	DC10	cerai	cerai	cerai	-	-	-	-

Syllable pattern: CVC

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
dan	DC1	dan	dan	dan	dan	dan	dan	dan
	DC2	dan	dan	dan	dan	dan	dan	dan
	DC3	dan	dan	dan	dan	dan	ada	ada
	DC4	dan	dan	dan	da	dan	dan	dan
	DC5	dan	dan	dan	dan	dan	dan	dan
	DC6	dengan	dan	-	-	-	-	-
	DC7	dan	dan	ban	dan	dan	dan	dan
	DC8	dan	dan	dan	dan	dan	dan	dan
	DC9	dan	ada	-	-	-	-	-
	DC10	dan	dan	dan	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
bin	DC1	bin	ben	bin	bin	bin	bin	bin
	DC2	bin	bin	bin	pin	pin	ben	bin
	DC3	bin	bin	bin	din	din	bin	dan
	DC4	bin	bin	pin	be'n	be'n	pin	be'n
	DC5	pin	bin	bin	bin	bin	bin	bin
	DC6	bin	bun	-	-	-	-	-
	DC7	bin	bin	ban	ban	ban	ban	ban
	DC8	bin	bing	bin	ban	bin	ban	ban
	DC9	pin	bin	-	-	-	-	-
	DC10	bin	bin	bin	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
wad	DC1	wad	wad	wad	wad	wad	wad	wad
	DC2	wad	wad	-	wap	wap	wap	wed
	DC3	man	wan	wan	wan	wan	wan	wan
	DC4	wad	wad	wad	wap	wad	kowap	wad
	DC5	we'n	we'p	we'p	wap	wad	wap	wan
	DC6	yad	wab	-	-	-	-	-
	DC7	wad	wa	wad	wan	wad	wan	wad
	DC8	ding	wan	wan	wad	wap	wam	wap
	DC9	wan	wan	-	-	-	-	-
	DC10	wap	wap	wap	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
pin	DC1	pin	pe'n	pin	pe'n	pin	pin	pe'n
	DC2	pin	pe'n	-	pin	pe'n	pin	pe'n
	DC3	pen	pin	pain	pin	pain	pe'n	pain
	DC4	bin	pin	pin	pin	pin	pin	pin
	DC5	pin	pin	pin	pin	pin	pin	pin
	DC6	bin	pin	-	-	-	-	-
	DC7	pin	pan	pin	pan	ban	pan	pin
	DC8	pe'n	pe'n	ping	pat	pan	pin	pe'n
	DC9	pin	pin	-	-	-	-	-
	DC10	pin	pun	pin	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
wap	DC1	wap	wap	wap	-	wap	wap	wap
	DC2	wap	wap	wap	wap	wep	wep	wap
	DC3	pun	wan	pin	wap	wip	put	put
	DC4	wap	wap	wap	wap	wap	wap	wap
	DC5	we'p	wap	wap	ap	wap	we'p	wap
	DC6	wap	wap	-	-	-	-	-
	DC7	wam	wapi	wap	wan	wap	wap	wap
	DC8	pom	wap	pang	wap	wap	uwap	wap
	DC9	wan	vin	-	-	-	-	-
	DC10	wap	wap	wap	-	-	-	-

Syllable pattern: V + CVC

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
umur	DC1	umur	umur	umur	umur	umur	umur	umur
	DC2	upur	-	-	umur	umur	umur	umur
	DC3	yuruh	umur	umur	umur	umur	umur	umur
	DC4	umur	umo	umur	umur	umur	umur	mmur
	DC5	umura	umur	mur	mur	umur	umur	umer
	DC6	-	umur	-	-	-	-	-
	DC7	amur	abur	pame'r	amur	amur	amar	amur
	DC8	umur	umur	yuror	umur	umur	umur	umur
	DC9	mur	mukar	-	-	-	-	-
	DC10	umir	umir	umir	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
abah	DC1	apa	abah	abah	abah	abah	abah	abah
	DC2	ada	abah	abah	ada	abah	abah	abah
	DC3	abah	abah	ada	abah	abah	abah	abah
	DC4	abak	abak	abu	abu	abu	ubah	ubah
	DC5	abah	abah	abah	abah	abah	abah	abah
	DC6	abah	abah	-	-	-	-	-
	DC7	abu	aha	abih	abur	abor	apu	abu
	DC8	adah	abah	baba	adah	adah	abang	abah
	DC9	ada	aha	-	-	-	-	-
	DC10	adah	adah	adah	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
ayat	DC1	ayat	ayat	ayat	ayat	ayat	ayat	ayat
	DC2	ya	ayah	ayah	ayah	ayat	ayat	ayah
	DC3	nyata	ngeti	tanya	tanya	tanya	nyata	nyata
	DC4	ayat	ayat	ayat	nyata	ayat	ayah	ayat
	DC5	-	ayah	ayah	ayah	ayah	ayah	ayah
	DC6	tayat	ayat	-	-	-	-	-
	DC7	anyat	ayah	ayi	ayi	anya	ayi	ayi
	DC8	ayah	ayap	ayam	ayah	ayah	ayam	ayap
	DC9	yang	atah	-	-	-	-	-
	DC10	ayah	ayat	ayat	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
upah	DC1	upah	upah	upa	upah	upah	upah	upa
	DC2	upah	upat	ubat	ubah	upah	upah	upah
	DC3	upah	upin	upah	upin	upah	upah	upah
	DC4	upah	upah	upon	upah	upah	upah	upah
	DC5	upah	upah	upah	upa	apah	ubah	apah
	DC6	upah	ubah	-	-	-	-	-
	DC7	apah	apa	aha	apah	abu	apa	apah
	DC8	puda	upur	upah	upah	pulau	upor	upor
	DC9	supaya	purah	-	-	-	-	-
	DC10	upah	upah	upah	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
ubah	DC1	ubah	uyubah	abah	ubah	ubah	ubah	uba
	DC2	upah	upah	ubah	ubat	ubah	ubat	ubah
	DC3	udah	uda	udah	udah	udah	udah	udah
	DC4	ubah	ubah	ubah	ubah	ubah	ubah	ubah
	DC5	ubah	ubah	ubah	yupah	ubah	ubah	ubah
	DC6	ubah	ubah	-	-	-	-	-
	DC7	abah	habo	abah	hadi	abu	abo	abu
	DC8	ubah	ubang	udang	ubah	subuh	ubang	ubat
	DC9	dulah	buah	-	-	-	-	-
	DC10	udah	udah	udah	-	-	-	-

Syllable pattern:CV + CV + CVV

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
senarai	DC1	senari	seri	senari	senarai	seri	seri	senarai
	DC2	sena	senari	senari	senarai	senarai	senarai	senaria
	DC3	surai	sundiri	serai	serai	suraini	surai	serani
	DC4	sinuri	sanari	seronon	serani	suruna	serina	seruni
	DC5	senarai	senarai	senarai	senarai	senarai	senarai	senarai
	DC6	-	senari	-	-	-	-	-
	DC7	sunar	senari	senna	senari	senari	seniri	sunirai
	DC8	senari	senarai	senarai	senarai	senarai	senarai	senarai
	DC9	sarik	saunna	-	-	-	-	-
	DC10	cenarai	senarai	senarai	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
serunai	DC1	seditnai	serunai	serunai	sunnya	seruni	seruni	serunyi
	DC2	seru	serunai	senarunai	serunai	serunai	serunai	serunai
	DC3	suri	serunai	surai	surai	suruni	surai	surani
	DC4	surana	suruna	seruna	suruni	suruni	suranu	seruna
	DC5	senurai	telunai	senurai	senurunai	serunai	serurai	serurai
	DC6	suni	senai	-	-	-	-	-
	DC7	suran	serun	seruna	serun	serunna	seri	seruni
	DC8	seruying	serunai	senunai	serunai	senurai	senurai	senunai
	DC9	sana	sini	-	-	-	-	-
	DC10	serunai	seruai	serunai	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
merayau	DC1	merayu	meriyau	merayu	merayu	merayu	merayu	merayu
	DC2	merayu	merayu	merayu	merayu	merayou	merayu	merayu
	DC3	haru	merapu	merayu	merayu	merayu	merayu	merayu
	DC4	merayu	merayu	merayu	merayau	merayu	merayu	merayu
	DC5	merayai	merayat	merayai	muraryai	merayai	merayai	menyai
	DC6	-	maru	-	-	-	-	-
	DC7	berai	mera	mera	meran	merayi	merinya	merinya
	DC8	merayo	merawai	merawai	merawai	merawai	merawai	merawai
	DC9	merayau	mayu	-	-	-	-	-
	DC10	merayau	merayu	merayu	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
belalai	DC1	belalai	baliai	belai	merarai	belai	berali	belali
	DC2	belali	belale	belari	belalai	belalai	belalai	belalai
	DC3	beli	nilai	membilai	membelai	membelai	membelai	membelai
	DC4	belali	belila	pelalan	belaling	belalang	belalang	bela
	DC5	belai	belalai	belangai	pelanglai	belarai	belalai	belalai
	DC6	meyi	belila	-	-	-	-	-
	DC7	beli	bela	bela	beli	bela	beli	beli
	DC8	belali	belarai	belalai	melalai	belalai	belalai	belalai

	DC9	belalah	bili	-	-	-	-	-
	DC10	belali	belalai	belalai	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
kemarau	DC1	kemeliau	kemariau	kemarau	kemariau	kemariau	kemariau	kemarau
	DC2	kemaru	kemaru	kemaru	kemaru	kemarou	kemaru	kemarau
	DC3	kemarau	kemurun	kemarau	kemarau	kemuruh	kumuruh	kemarau
	DC4	kemurahan	kemurangan	kemurang	kemuraran	kemaru	kemurangan	kemurang
	DC5	kemarai	kemarai	kemarai	kemarai	kemarai	kemarai	pemarai
	DC6	kamar	kerok	-	-	-	-	-
	DC7	kemar	kemar	kemar	kemaru	kemmari	kemmaru	pemmaru
	DC8	kemaran	kemeran	kemarur	kemarai	kemerai	kemerai	kemurai
	DC9	kemuru	umur	-	-	-	-	-
	DC10	kemaru	kemaru	kemaruai	-	-	-	-

Syllable pattern: CV + CV + CCVV

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
perangai	DC1	perangai	perangai	perangai	perangai	pepancing	pelbagai	perangai
	DC2	berngai	berangkat	berangkat	peranai	peraogai	peragai	peran_gai
	DC3	pembarai	peringi	meragai	perangai	perbagai	penyayang	perana
	DC4	peranai	peranai	peranai	peranai	peranai	peranai	peranai
	DC5	peregai	perangkai	peranggai	pelangkai	peranggai	perangkai	perangkai
	DC6	manggi	pebangai	-	-	-	-	-
	DC7	purana	perna	penag	benargi	beran_gi	periga	meringga
	DC8	peragi	perenggan	peranggai	perana	peragai	peragai	peragai
	DC9	pergi	pagi	-	-	-	-	-
	DC10	pernagai	peranggai	peranggai	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
jerangau	DC1	-	-	jerangau	jeralan	jerangou	jerangau	jerangau
	DC2	jingpunan	jiran	jeragu	jerangkut	jerangut	jeranggut	jeragau
	DC3	penjuru	jeragul	meracau	merajuk	jemuruh	jaruh	jahuru
	DC4	jerugah	jeragul	jerang	jengkit	jaru	jerung	ejerit
	DC5	jerangku	Jerangkan	jerangkut	jerangkut	jeranggai	jerangkoiyuk	jerangkai
	DC6	mayut	jengu	-	-	-	-	-
	DC7	jerag	jerug	jerag	jegi	jerinka	jerin	jiring
	DC8	jerawat	jeranjai	jeranggai	jenawai	jeragai	geregai	jeragu
	DC9	jira	jumur	-	-	-	-	-
	DC10	jeragun	ja_anggu	jeragan	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
meringai	DC1	semak	merangai	merangai	merangai	merangai	merangi	merangai
	DC2	merinang	merinak	merinai	merikai	merigai	merin_gai	merigai
	DC3	penjarai	meringi	melihat	meringal	memberi	memerah	mena_erai
	DC4	merinaga	merinagi	meranggi	merogi	mere'ngang	harugikan	pere'ni
	DC5	merigai	merungkai	meringgai	meyingkai	meringgai	meringkai	meringkai
	DC6	ramai	meramai	-	-	-	-	-
	DC7	berig	merti	meni	mereni	mernyani	meringga	mering
	DC8	merangji	merenjai	merigai	melanjai	merigau	meragai	merigau
	DC9	merinningkan	mennya	-	-	-	-	-
	DC10	meringgan	meringgai	meringga	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
melangai	DC1	melangai	melangai	melangai	langai	melangai	melangi	melangai
	DC2	melangai	melainak	melaning	melakai	melangai	melan_gai	melan_gai
	DC3	melangai	mengilai	-	melangai	membelai	membelai	membelai
	DC4	melagi	melagi	menanggi	melaye	melangga	melanggar	melangkit
	DC5	melanggai	melangkai	melanggai	melanggai	melanggai	melangkai	menangkai
	DC6	manggai	memangai	-	-	-	-	-

	DC7	mengai	malag	mele't	melingga	menami	melangga	melingka
	DC8	melangji	melajai	melanai	melagai	belajar	melagai	melanai
	DC9	melalur	mengkejaan	-	-	-	-	-
	DC10	melanggi	melanggai	melan_gi	-	-	-	-

Syllable pattern: CVC + CV + CVC

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
kempunan	DC1	kepunyaan	kampungan	kempunan	kempunan	kampunan	kampunan	kampunan
	DC2	-	kempulan	kempunan	kempulan	kempunan	kempunan	kepunan
	DC3	kepunai	kepuntan	kempunan	kumputan	kepulitan	kempunan	kempunan
	DC4	kepuni	kumpu_a_nan	kepunyaan	kemumpanu	kepunyaan	kepuyaan	kempumaan
	DC5	kepungan	kempunan	kepunang	kepula	kempunan	kempunang	kepunan
	DC6	puna	kemaye'n	-	-	-	-	-
	DC7	kempun	kampun	kampur	kempurna	kampur	kampur	kembuna
	DC8	kempunang	kempuna	kempuna	kempunah	kamputan	kempunan	gempunan
	DC9	pernya	kempiuna	-	-	-	-	-
	DC10	kempuan	kempuan	kempuan	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
jemputan	DC1	jemputan	jamputan	jemputan	jamputan	jembutan	jemputan	jemputan
	DC2	jemputan	jimputan	jempunan	jemputan	jempunan	jemputan	jeputan
	DC3	jambatan	jemputan	jambatan	jembatan	jambatan	jembatan	jembatan
	DC4	jemputan	jemputan	jemputan	jemputan	juputan	jepuntan	jemputan
	DC5	jemputang	jempuntang	cemutang	cemputan	jepuntan	jepuntang	jeputang
	DC6	jemput	jemputan	-	-	-	-	-
	DC7	jemtut	jemtang	jamputan	jembutong	jambor	jempa	jembutah
	DC8	jemputan	jemputtan	jemputtan	jemputtan	jemputtan	jemputtan	jemputtan
	DC9	jemputtan	jemputan	-	-	-	-	-
	DC10	jemputan	jemputan	jemputan	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
kumpulan	DC1	kumpulan	kumpulan	kumpulan	kumpulan	kumpulan	kumpulan	kumpulan
	DC2	kumpulan	kumpulan	kumpulan	kempulan	kumpulan	kumpulan	kumpulan
	DC3	kupulangan	lupaan	pulangan	kepuloi	kumpunan	kumpulang	kempulang
	DC4	kupulan	kupulan	kumpulan	kumpulan	kupulan	kupulan	kumpulan
	DC5	kumpulang	kampunah	kumpulang	kumpulan	kumpulan	kumbulan	kumpulan
	DC6	kempulan	kempulen	-	-	-	-	-
	DC7	kumpulong	kempun	kempul	kembulong	kampur	kempola	kempulan
	DC8	kampunglan	tempatlah	kepulau	kemula	punglan	kemputtan	jemputlan
	DC9	kumpulan	kumpulan	-	-	-	-	-
	DC10	kempulan	kumpulan	kumpulan	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
cendawan	DC1	cendawan	cendawan	cendawan	cendawan	cendawan	cendawan	cendawan
	DC2	cendawan	cendawan	cendawan	cendawan	cendawan	cendawan	cendawan
	DC3	cewangan	cendawan	cendawan	cendawan	cendawan	cendawan	cendawan
	DC4	cendawan	cendawan	cendawan	cindawan	cendawan	cendawan	cendawan
	DC5	cedawan	cedawan	dedawan	cedawan	cedawan	cedawan	cedawan
	DC6	cendan	cedawan	-	-	-	-	-
	DC7	canwan	hantu	sandawa	sendawan	kandawan	sandawan	jengdawan
	DC8	sedawan	cedawa	cedawai	cendawan	cendawan	jembawang	cendawan
	DC9	cendwan	sindiawan	-	-	-	-	-
	DC10	cendawan	cendawan	cendawan	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
maklumat	DC1	muluat	maklumat	maklumat	maklumat	maklumat	maklumat	maklumat
	DC2	makulmat	mak_l_mat	makulmat	makelmat	makulmat	mak_umat	maklumat
	DC3	maklumat	meluan	maklumat	maklumat	maklumat	maklumat	maklumat
	DC4	malumat	maklumat	maklumat	aknumat	maklumat	maklumat	maklumat
	DC5	malumat	maklumat	maklumat	maklumat	maklumat	maklumat	maklumat

	DC6	mamut	melut	-	-	-	-	-
	DC7	-	kalum	maklong	maglong	makwan	kantum	maklong
	DC8	melumat	kelumat	kelumat	kelumat	kelua	malumat	kelubak
	DC9	perlautan	mengkeliatan	-	-	-	-	-
	DC10	makulmat	makkumat	mak_umat	-	-	-	-

Syllable pattern: CV + CV + CVCC

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
binatang	DC1	binatang	banyatan	binatang	pinatang	binatang	binatang	binatang
	DC2	pindah	bintang	bintang	bintai	minatang	binatang	binatang
	DC3	metang	baintang	-	binotong	bintang	binatang	bintang
	DC4	binatang	binatang	binatang	natang	binatang	binatang	binatang
	DC5	pintunatang	binatang	bintang	bintangan	bintangan	binatang	bintangan
	DC6	binatan	binatan	-	-	-	-	-
	DC7	bintagik	binak	bintah	binnate'ng	binatang	binatang	binatang
	DC8	binatang	binatang	binatang	benatai	binatang	binatang	binatang
	DC9	pentoweltan	bintang	-	-	-	-	-
	DC10	binatang	binatang	bintang	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
terowong	DC1	terowonat	terowang	terowong	teroyong	terowong	terowong	terowong
	DC2	terbong	terowok	terwo	terowok	terowong	terwo	terowong
	DC3	tuporang	teborong	tumoro	tumoro	-	terowong	terowong
	DC4	towong	corong	terong	torong	-	torong	torong
	DC5	teroron	terongrong	terongron	terongrong	terowong	terongrong	terongrong
	DC6	tong	temborong	-	-	-	-	-
	DC7	tewan	-	tero	terawe	teroyang	terowo	terowong
	DC8	terowang	terojang	teronggor	teruwai	terenjang	terowang	terowo
	DC9	terolong	ceropotan	-	-	-	-	-
	DC10	terowongan	terowang	terowongan	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
belalang	DC1	belalang	balang	belang	belalang	belalang	balang	belalang
	DC2	belakang	belalang	belalang	belalai	belalai	belalang	pelalang
	DC3	balang	balang	bilang	pulan	belalang	bilang	bilang
	DC4	-	lalang	belaling	belalang	belalang	belalang	belalang
	DC5	belalang	belanglang	belalang	belangkan	belalang	pelangkan	pelalang
	DC6	belang	belang	-	-	-	-	-
	DC7	belagak	berula	belag	berlanang	belang	belakang	belangka
	DC8	belalang	belalang	belalang	belalang	belalang	belalang	belalang
	DC9	berlawan	beedi	-	-	-	-	-
	DC10	belalang	belanglang	belale'ng	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
belakang	DC1	belakang	belangngan	belangkang	belakang	belakang	belakang	belakang
	DC2	belakin	belakon	belakan	belakang	belakon	belakon	belakan
	DC3	melakan	belakang	melakang	belakang	belakang	belang	pelakang
	DC4	blakang	belakang	belakang	blakang	blakang	blakang	blakang
	DC5	pelangkan	belakan	pelangkan	belangkan	belangkang	pelangkan	pelangkan
	DC6	benakan	belakan	-	-	-	-	-
	DC7	belagan	belag	belikat	belakan	belakan	belakan	pelakang
	DC8	belaku	belakang	belakang	belaku	belakang	belakang	belakang
	DC9	bilakan	belekat	-	-	-	-	-
	DC10	belakang	belakang	belakang	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
meradang	DC1	meradang	meradang	merendang	meradang	meradang	meradang	meradang
	DC2	berkat	merana	meladang	meratai	meradai	meradang	merredang
	DC3	meradang	medang	meradak	merana	menaran	meriang	pengwarang
	DC4	meradik	meradik	merendah	merendak	meradak	harandah	merendak
	DC5	merandang	merandang	merada	mere'ndang	merangdang	merendang	perandang
	DC6	nadang	meradan	-	-	-	-	-
	DC7	merag	terna	merag	meradi	meragi	meraga	merangda
	DC8	merada	meradang	merandang	meradai	meladang	meradang	merada
	DC9	mengkadukan	mengedak	-	-	-	-	-
	DC10	meradangkan	meradangan	meradangan	-	-	-	-

Syllable Pattern: CV+CVC+CVCC

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
penumpang	DC1	penumpang	penyumpang	penumpang	penumpang	penumpang	penumpang	penumpang
	DC2	pen_um_pung	penumpat	penumpat	penumpat	penumpat	penumpat	penumpang
	DC3	penumpang	penumpang	penumpang	penumpang	penumpang	pembilang	penumpang
	DC4	penumpang	penumpang	penumpang	unupang	menumpang	penupang	menumpang
	DC5	penumpak	menumpak	menumpak	mere'ndang	pennumpak	penumpak	benumpak
	DC6	pengang	menupan	-	-	-	-	-
	DC7	benampan	penna	pennakpan	pennapan	penampung	penampang	penumpon
	DC8	penajang	penampai	pemangpai	pemanpo	penampai	penapai	menangpai
	DC9	pelagungan	peniaga	-	-	-	-	-
	DC10	penyumpa	pemumpaan	pinmupangan	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
memancing	DC1	memancing	memancing	memancing	menyicang	memancing	memancing	memancing
	DC2	meancing	memanjat	memanjat	memmacat	memanjat	memanjat	memancing
	DC3	mencari	mincing	menciri	melaci	mengaji	melaici	menaci
	DC4	memancing	memacing	memancing	memancing	memancing	memancing	menancing
	DC5	memancing	memankit	memancing	menamcit	memangcing	memangcing	memancing
	DC6	benacang	melacan	-	-	-	-	-
	DC7	menagin	menag	meman	menari	menanggi	mennaki	menanggi
	DC8	menese'k	melencang	mecencing	menecang	menanjai	memacing	memacing
	DC9	mencari_ing	mencebasan	-	-	-	-	-
	DC10	mem_an_cing	memaicinnin	memacing	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
menendang	DC1	merendang	mejandang	menengdang	menindang	menandang	menendang	menendang
	DC2	meninda	meneda	memninda	menindai	menendak	menendang	menendang
	DC3	memidak	mendetang	mengidang	meladang	merendang	mengidam	mengidah
	DC4	menendang	menendang	mendang	menendang	menandang	nendang	menandang
	DC5	menendak	menendak	menendak	menendak	menendak	menendak	menendak
	DC6	mendang	medang	-	-	-	-	-
	DC7	menanding	neda	menang	menida	menanggi	meninggah	menangda
	DC8	menedang	menendang	menengdang	menadai	menandai	menadang	menadang
	DC9	berendaraan	menjajan	-	-	-	-	-
	DC10	menindangan	memindangan	memandawan	-	-	-	-

Target	Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7

melenting	DC1	melenting	melenting	melenting	melintang	melanting	melenting	melenting
	DC2	melinting	melintit	melintik	malintai	melentik	melintik	melinting
	DC3	menanti	menentang	melati	melintang	melentang	melentang	memmati
	DC4	meletihkan	meletihkan	meletih	meletihkan	melatik	melenting	meletihkan
	DC5	welenting	melentik	melentik	melentik	melengtik	melengtik	melingtik
	DC6	tengen	melatang	-	-	-	-	-
	DC7	melangtik	menag	melang	menangtang	melangting	melangting	mele'tang
	DC8	melatih	melatang	melantai	melutai	melatil	melating	melatik
	DC9	menapolan	meEDatan	-	-	-	-	-
	DC10	melantas	melantingan	melantir	-	-	-	-

Target		**Session 1**	**Session 2**	**Session 3**	**Session 4**	**Session 5**	**Session 6**	**Session 7**
kecundang	DC1	kecunju	kecundang	kecundang	kecundang	kecunduang	kecundang	kecundang
	DC2	kecundang	kecundang	kecundan	kecundai	kecundan	kecundang	kecundang
	DC3	kejudang	kecundang	kucundang	kecundang	kucawan	kuncang	kecundang
	DC4	kecuadan	kucitikan	kenacang	e'cunanan	kecundan	kecindung	kecunjingkan
	DC5	kecudang	kecundah	kecunda	kecundang	kecundang	kecungdang	kecudang
	DC6	sedang	kedang	-	-	-	-	-
	DC7	kesunnga	sunda	kenada	kekunding	kekanggi	kelongda	kelongda
	DC8	kecundan	kecedang	kecondok	kecendag	pencudang	kecandang	kecedang
	DC9	belatan	kicapdan	-	-	-	-	-
	DC10	kecundangu	kecudangan	kecundangan	-	-	-	-

Syllable pattern:V+ CVCC

Target		**Session 1**	**Session 2**	**Session 3**	**Session 4**	**Session 5**	**Session 6**	**Session 7**
abang	DC1	abang	abang	abang	abang	abang	abang	abang
	DC2	atuk	amang	abang	adang	abang	abang	abang
	DC3	adang	abang	abang	abang	abang	abah	abang
	DC4	abang	abang	abang	abang	abang	abang	abang
	DC5	abang	abang	ubangnga	abang	abang	abang	abang
	DC6	bengugeng	abang	-	-	-	-	-
	DC7	angan	adang	abing	api	abo	apa	abu
	DC8	adang	abang	abong	abang	abang	abang	abang
	DC9	abang	bang	-	-	-	-	-
	DC10	adangu	abangan	adangan	-	-	-	-

Target		**Session 1**	**Session 2**	**Session 3**	**Session 4**	**Session 5**	**Session 6**	**Session 7**
orang	DC1	orang	orang	orang	orang	orang	orang	orang
	DC2	orang	orang	orang	orang	upah	orang	orang
	DC3	orang	orang	orang	orang	orang	orang	orang
	DC4	orang	orang	orang	orang	orang	orang	orang
	DC5	urang	orang	orang	orang	orang	orang	orang
	DC6	orang	orang	-	-	-	-	-
	DC7	ore'n	ore'n	ore'n	oren	oren	oren	oren
	DC8	orang	orang	orang	orang	orang	orang	orang
	DC9	orang	orang	-	-	-	-	-
	DC10	orang	orang	orang	-	-	-	-

Target		**Session 1**	**Session 2**	**Session 3**	**Session 4**	**Session 5**	**Session 6**	**Session 7**
ulang	DC1	ulang	ulang	ulang	ulang	ulang	ulang	ulang
	DC2	ular	ulang	ulat	ulat	ulat	ulat	ulat
	DC3	ulang	ular	ulang	ulang	ulang	ulang	ulang

	DC4	ulanan	ulang	ulalan	ulang	ulang	ulang	ulang
	DC5	ulang	ulang	ulang	ulah	ula	ulang	ulang
	DC6	lungu	olang	-	-	-	-	-
	DC7	ulang	ale'k	ulag	aling	lulang	ala	aling
	DC8	ula	ulang	ule'ng	ula	ular	ula	ula
	DC9	lalu	lugang	-	-	-	-	-
	DC10	undang	ular	ulang	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
udang	DC1	udang	udang	udang	udang	udang	udang	udang
	DC2	udang	ubat	ubat	ubat	ubat	ubat	udang
	DC3	udang	udang	udang	udang	udang	udang	udang
	DC4	udang	udang	udang	udang	udang	udang	udang
	DC5	udang	udang	udang	udang	udang	udang	udang
	DC6	udang	udang	-	-	-	-	-
	DC7	adik	ada	udah	adi	luga	ada	ada
	DC8	udang	udang	udang	udang	udang	udang	udang
	DC9	lulukan	ludar	-	-	-	-	-
	DC10	undang	undang	undang	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
inang	DC1	inang	inang	inang	inding	inang	inang	ingin
	DC2	inak	inak	ina	inak	inak	inak	inang
	DC3	hilang	inan	indang	indah	undang	hilang	hilang
	DC4	inna	inan	inna	inang	innan	innan	inah
	DC5	innak	inagk	inga	inngah	innak	inak	innak
	DC6	angen	ani	-	-	-	-	-
	DC7	innag	anya	indang	nina	cintang	ingnga	inan
	DC8	inang	ine'ng	inai	innai	inai	inai	inai
	DC9	aiding	ingat	-	-	-	-	-
	DC10	inyang	inangai	miar	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
rumah	DC1	rumah	rumah	rumah	rumah	rumah	rumah	rumah
	DC2	rumah	rumah	rumah	rumah	rumah	rumah	rumah
	DC3	rumah	rumah	rumah	rumah	rumah	rumah	rumah
	DC4	rummah	rumah	rumah	rumar	rumah	rumah	rumah
	DC5	rumah	umah	umah	umah	rumah	umah	rumah
	DC6	orum	rumah	-	-	-	-	-
	DC7	rummah	rumna	rumah	rumah	ruman	rummah	rima
	DC8	rumah	rumah	rumah	rumah	rumah	rumah	rumah
	DC9	rumah	rumah	-	-	-	-	-
	DC10	rumah	rumah	rumah	-	-	-	-

Syllable pattern: CV+CVC/+CV

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
betul	DC1	betul	betul	betul	betul	betul	betul	betul
	DC2	betul	betul	belur	belur	betul	betul	betul
	DC3	betul	betul	betul	betul	betul	betul	betul
	DC4	betul	betul	betul	betul	betul	betul	betul
	DC5	belum	botol	botol	betol	betul	beltul	botol
	DC6	botol	butul	-	-	-	-	-
	DC7	betuk	belu	batak	botol	betal	betal	betul
	DC8	betul	betul	betel	betul	betul	betul	betul
	DC9	beli	buzutan	-	-	-	-	-
	DC10	betul	betul	betul	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
makan	DC1	makan	makan	makan	makan	makan	makan	makan
	DC2	makan	makan	makan	makan	makan	makan	makan
	DC3	makan	makan	makan	makan	makan	makan	kan
	DC4	makan	makan	makan	makan	makan	makan	makan
	DC5	makan	makan	makan	makan	makan	makan	makan
	DC6	makan	makan	-	-	-	-	-
	DC7	mekan	mamkan	makan	makan	makan	makan	makan
	DC8	manan	makan	makan	makan	makan	makan	makan
	DC9	makanan	makan	-	-	-	-	-
	DC10	makan	makan	makan	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
bijak	DC1	bijak	bijak	biji	bijak	bijak	bijak	bijak
	DC2	biji	bijak	pijak	bijak	centik	bijak	bijak
	DC3	manji	bijik	bijik	bijak	bijak	bijik	majik
	DC4	bijak	bijak	bijak	bijak	ijak	bijak	bijak
	DC5	bijik	bijak	bijak	bijak	bijak	bijak	bijak
	DC6	bijak	bijak	-	-	-	-	-
	DC7	bijak	bijak	bekaih	bejaka	bijang	beki	bijik
	DC8	bijak	bijak	bijak	bijak	bijak	bijak	bijak
	DC9	bijak	bijak	-	-	-	-	-
	DC10	bidak	bijak	bijak	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
pergi	DC1	pergi	perangai	peringi	pergi	pergi	pergi	pergi
	DC2	pergi	pergi	pergi	pergi	pergi	pergi	pergi
	DC3	pergi	pergi	pergi	pergi	pergi	pergi	pergi
	DC4	piring	pegi	pegi	pegi	pegi	pegi	pegi
	DC5	bergi	pergi	pergi	pergi	bergi	pergi	pergi
	DC6	perji	pegi	-	-	-	-	-
	DC7	perga	bera	perag	pegi	pargi	pering	pargi
	DC8	bedi	pergi	pergi	pergi	pergi	pergi	pergi
	DC9	pergi	pergi	-	-	-	-	-
	DC10	pergi	pergi	pergi	-	-	-	-

Syllable Pattern: CV+CVCC

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
sayang	DC1	sayang	sayang	sayang	sayang	sayang	sayang	sayang
	DC2	sayur	saya	sayap	sayur	sayur	sayur	sayang
	DC3	sayang	sayang	sayang	sayang	sayang	sayang	cayang
	DC4	sayang	sayang	layang	sayang	sarang	saya	sayang
	DC5	sayang	sayang	sayang	sayang	sayang	sayang	sayang
	DC6	-	senge	-	-	-	-	-
	DC7	sayang	sanya	saya	sawan	sayang	seyi	saya
	DC8	sayang	sayang	sayang	sayang	sayang	sungai	sanya
	DC9	sayang	saya	-	-	-	-	-
	DC10	sayang	sayang	sayang	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
bawang	DC1	bawang	bawang	bawang	bawang	bawang	bawang	bawang
	DC2	-	bawang	bawat	bawang	bawang	bawa	bawang
	DC3	bawang	bawang	bawang	bawang	bawang	bawa	bawah
	DC4	bawang	bawang	bawang	bawang	barang	bawang	awing
	DC5	bewang	bawang	bawang	bawang	bawang	bawang	bawang
	DC6	badan	bawang	-	-	-	-	-
	DC7	payang	bewa	bawa	bawang	bayang	bawang	bawang
	DC8	bewang	bawang	bawang	bawang	bawang	bawang	bawa

	DC9	berlaku	bawang	-	-	-	-	-
	DC10	bewangan	bawangan	bawang	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
sarang	DC1	serang	sarang	sarang	sarang	serang	sarang	sarang
	DC2	saring	sarang	sarang	sarung	sarang	sarang	sarang
	DC3	seram	serang	seorang	sarang	sarang	sarang	sarang
	DC4	sarang	sarang	sarang	sarang	sarang	sarang	sarang
	DC5	saran	sarang	sarang	sarang	sarang	sarang	sarang
	DC6	-	senge	-	-	-	-	-
	DC7	sarang	sarang	sarang	sare'ng	sarang	sari	sare'ng
	DC8	sarang	serang	sarang	serang	sarang	sayur	sara
	DC9	sukan	saka	-	-	-	-	-
	DC10	sarangan	sarangan	sarang	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
barang	DC1	barang	barang	barang	barang	barang	barang	barang
	DC2	barang	barang	barang	barang	barang	barang	barang
	DC3	barang	birang	barang	barang	baring	barang	barang
	DC4	barang	barang	barang	parang	burung	barang	barang
	DC5	barang	barang	barang	barang	bawang	barang	barang
	DC6	badan	barang	barang	barang	bawang	barang	barang
	DC7	berna	barang	barang	bare'ng	baring	baring	bare'ng
	DC8	bewat	barang	berang	barang	barang	barang	barang
	DC9	berkan	barkar	-	-	-	-	-
	DC10	barangan	barangan	barangan	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
baring	DC1	baring	baring	baring	baring	baling	baling	baring
	DC2	baring	baring	baring	baring	baring	baring	baring
	DC3	birang	birang	barang	berin	bari	beriang	birang
	DC4	baring	baring	baring	baring	baring	aring	baring
	DC5	baring	baring	baring	baring	baring	baring	baring
	DC6	barang	baring	-	-	-	-	-
	DC7	bernag	berang	berag	berni	berning	bering	berni
	DC8	bering	badang	bering	bering	bari	bering	bari
	DC9	pernyan	berni	-	-	-	-	-
	DC10	baringan	baringan	baringan	-	-	-	-

Syllabe Pattern: CV+CVC/CVC+CV

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
pantai	DC1	bantai	banting	pantai	pantai	pantai	pantai	pantai
	DC2	mantai	bantai	bantal	pantai	bantai	bantai	pantai
	DC3	pita	tempat	pita	pintak	petai	betani	petani
	DC4	pantai	pantai	pantai	pantai	pantai	pantai	pantai
	DC5	pantai	pantai	pantai	pantai	pantai	pantai	pantai
	DC6	pantai	batai	-	-	-	-	-
	DC7	pantan	pentan	panta	pentang	bentang	banta	bantang
	DC8	penak	perti	penting	penting	pantai	pantai	peta
	DC9	pententu	parta	-	-	-	-	-
	DC10	pantai	pantai	pantai	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
sangat	DC1	sangat	sangat	sangat	sangat	sangat	sangat	sangat
	DC2	sangan	sangat	sanngat	sanngat	sangat	semgan	sangat
	DC3	santi	sengat	sorang	sati	cere'ti	-	se'hat
	DC4	angat	sanat	sanat	sanat	sanat	sanat	sanat
	DC5	sarang	sangkat	sangkat	sangkat	sangkat	sangga	sangkat
	DC6	sangat	sangat	-	-	-	-	-

	DC7	san_gan	san_gi	sangkan	sangkan	cangan	sandang	san_gan
	DC8	senjak	sanjung	sangjak	sunjak	saja	sambat	sunyi
	DC9	satkan	satgega	-	-	-	-	-
	DC10	sanggat	sangat	sangat	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
jangan	DC1	jangan	jangan	jangan	jangan	jangan	jangan	jangan
	DC2	jangan	jang an	jangan	jangan	jangan	jangan	jangan
	DC3	ajang	jengan	jalan	jangan	jangan	jangan	jangan
	DC4	janan	janan	janan	janan	janan	janan	janan
	DC5	jangkan	jangkat	jangkat	jangkat	jangkat	jangga	jangkat
	DC6	jajan	jangan	-	-	-	-	-
	DC7	jantan	jakit	jantak	jakin	janggan	jangkan	jangkan
	DC8	janjan	jengja	jemput	jenang	jaga	jangjan	jenja
	DC9	jemputtan	juga	-	-	-	-	-
	DC10	jangkat	jangat	jangat	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
pandai	DC1	pandai	pandai	pandai	pandai	pandai	pandai	pandai
	DC2	banding	pandai	pandai	pandai	bantai	pandai	pandai
	DC3	pedi	pina	pidai	pendah	bidai	pelan	pedai
	DC4	pandai	pandai	andai	pandai	pandai	pandai	pandai
	DC5	pantai	bantai	pandai	pantai	pantai	pantai	pantai
	DC6	pandai	penagai	-	-	-	-	-
	DC7	pandi	-	panti	pading	banding	pandan	pandan
	DC8	penggan	pendang	pedang	pedang	pandang	pandan	pandang
	DC9	padanya	padah	-	-	-	-	-
	DC10	pandi	pandi	pandi	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
pangsa	DC1	penjasa	pingas	penjasa	penjasa	penjasa	penjasa	penjasa
	DC2	panngas	pansa	pansat	pangsa	pangsa	pengsan	bangsa
	DC3	panas	pisah	pesa	pisah	pasang	pisang	pisah
	DC4	pangsa	pasang	angsa	pangsa	pangsa	pangsa	pangsa
	DC5	mangsa	bangsa	bangsa	pangsa	pangsa	pangsa	pangsa
	DC6	pasa	pemasa	-	-	-	-	-
	DC7	pan_gei	pangsa	pangas	bangsa	pangsang	bangsan	pengsan
	DC8	pensa	pansa	pengsa	pangsa	pangsa	pangsa	pangsu
	DC9	perdahukun	pakas	-	-	-	-	-
	DC10	panggas	panggas	pan_gas	-	-	-	-

Syllable Pattern: CVC+CVC

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
keldai	DC1	keldai	keldai	keldai	kelaidi	keladi	kelaidi	kelaidi
	DC2	keldai	kedal	kelat	keldai	keldai	keldai	keldai
	DC3	keldai	kedai	kaldai	keldal	keldai	keldal	kaldal
	DC4	kelda	keldai	kedai	keldai	keldai	keldai	keldai
	DC5	keldai	keldai	keldai	kedai	keldai	keldai	keldai
	DC6	kaldai	kebagai	-	-	-	-	-
	DC7	kaldi	keldat	keeda	kali	keldang	keldan	kelit
	DC8	kede'k	keldak	keldang	ledak	kandang	teldak	kelu
	DC9	keldip	keda	-	-	-	-	-
	DC10	keladi	keldi	keldi	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
hampir	DC1	hampir	hampir	hampir	hampir	hampir	hampir	hampir
	DC2	hampir	hampir	hampir	hampir	hampir	hampir	hampir
	DC3	barang	hampa	harapan	harapan	hampar	sarapan	harapan
	DC4	harap	harap	harap	harap	harap	harap	harap

	DC5	harpit	harmpit	harmpit	harmpat	harmpit	harmpit	harmpit
	DC6	hapir	hapai	-	-	-	-	-
	DC7	hampar	hambir	hampur	mapir	hambur	hampur	hambor
	DC8	lapor	bapa	hampir	hampir	hampir	hampir	hampir
	DC9	he'mper	harpa	-	-	-	-	-
	DC10	himpi	hampiran	himpir	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
pernah	DC1	per_ah	perana	peranah	perana	peranah	peranak	perana
	DC2	pernah	pertah	perrah	berah	pernah	perah	bernah
	DC3	pernah	pernah	pera	pernah	pernah	pernh	pernah
	DC4	penah	penah	enah	penah	enah	penah	penah
	DC5	perangk	penar	penar	pernar	pernah	pernah	penah
	DC6	pahnah	penah	-	-	-	-	-
	DC7	perka	pernaha	pernah	pernah	perning	pernah	pernah
	DC8	damtai	penang	pernat	pernat	pernah	pernah	pernah
	DC9	pernya	cineryah	-	-	-	-	-
	DC10	peranah	peranai	pirham	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
cantik	DC1	cantik	cantik	cantik	cantik	cantik	cantik	cantik
	DC2	cantik	cantik	cantik	cangtik	cantik	cantik	cantik
	DC3	cantik	cantik	cantik	cantik	cantik	cantik	cantik
	DC4	cantik	cantik	natik	cantik	cantik	cantik	cantik
	DC5	cantik	cantik	cantik	cantik	cantik	cantik	cantik
	DC6	cantik	cantik	-	-	-	-	-
	DC7	cantik	kante'	kantik	sante'ng	canting	santik	nantan
	DC8	cantik	cantik	cantik	cendik	cindak	cantik	cantik
	DC9	catan	catak	-	-	-	-	-
	DC10	cantik	cantik	cantik	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
hampar	DC1	hampar	hampar	hampar	hampir	hampir	hampir	hampar
	DC2	hampar	hapar	hambar	hampar	hampir	hampir	hampir
	DC3	hampar	hampar	hampar	harapan	harapan	harapan	harapan
	DC4	haram	haram	haram	aram	ampar	hampar	hampar
	DC5	harabaq	hampart	harmpart	harmpart	harmpart	hampat	harmpat
	DC6	hampar	hampar	-	-	-	-	-
	DC7	hampan	hampan	hamra	hempar	hampur	hampur	hampor
	DC8	he'nper	hemrur	rumpar	hampir	hampir	hampa	hampir
	DC9	hamarkan	parpar	-	-	-	-	-
	DC10	hampar	himpir	hampir	-	-	-	-

Syllable Pattern: CVC+CVC

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
kangkung	DC1	kangkung	kanjikun	kangkun	kangkun	kangkun	kangkun	kangkun
	DC2	-	kamkuk	kamkur	kankun	kangkung	kampung	gangkung
	DC3	kerang	kerang	kulang	kangkung	-	kangkung	kelang-kelang
	DC4	kuka	kukang	kuka	kanggu	kuka	kuka	kungga
	DC5	kankun	kankung	kankun	kankun	kankun	kankun	kankun
	DC6	kankin	kankun	-	-	-	-	-
	DC7	gantung	ngantuk	kankun	kankun	kangkur	gangkur	kangkor
	DC8	kerkur	kenkun	kantong	kenkun	penkun	kankun	kamkung
	DC9	celuan	kura-kura	-	-	-	-	-
	DC10	kan_gungan	kunyinkuan	kangankuan	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
jengking	DC1	jakin	angkin	jangkit	jingkang	jengkit	jangkit	jingkit

	DC2	jengkin	jenkit	jenakit	jinkit	jekit	jengkit	jengkit
	DC3	kenji	jerang	jengking	jajiga	jengkai	jaji	jangjitan
	DC4	jingka	dengki	jangkit	japjig	jakit	jengki	jangkit
	DC5	jenkit	jenkit	jengkit	jengkit	jengkit	jengkit	cengkit
	DC6	jakan	jantik	-	-	-	-	-
	DC7	janggan	jangkuk	jentik	jakin	gatang	je'ngkur	jangkor
	DC8	jantan	jemking	jengking	jenkak	jingking	jenkik	jemking
	DC9	jankan	kerja	-	-	-	-	-
	DC10	jengkian	jenginkian	jengkian	-	-	-	-
Target		**Session 1**	**Session 2**	**Session 3**	**Session 4**	**Session 5**	**Session 6**	**Session 7**
pinggang	DC1	pinggang	-	panjang	pingngang	panjang	pinngan	pingang
	DC2	penkan	bintang	binkah	pingan	pinggang	pinggah	pinggang
	DC3	penji	pirang	pejan	belalang	pisah	bilang	pisah
	DC4	pinggan	pinggan	pinggang	pinggan	pinggan	pinggan	pinggan
	DC5	pingkat	pinggat	pinggan	pingban	pinggan	pinggan	pinggan
	DC6	ajan	penjan	-	-	-	-	-
	DC7	pinggan	pinjang	penja	pinggan	pangkur	panggur	pangkor
	DC8	pergan	perjang	pendang	penjak	penja	pinjam	pinjam
	DC9	pergi	pergi	-	-	-	-	-
	DC10	pinggian	penggian	pinggian	-	-	-	-

Target		**Session 1**	**Session 2**	**Session 3**	**Session 4**	**Session 5**	**Session 6**	**Session 7**
bengkung	DC1	bemungkun	panjikun	angkun	bangun	bangkun	mengkung	mangkun
	DC2	banngun	bangkut	bankut	pengun	bengkut	pangkut	pengkut
	DC3	becakung	bangkung	melentang	belakon	pelakang	bengkung	be'ngkok
	DC4	buku	bangku	bangku	bangkuk	bangkit	angku	bangkuk
	DC5	wankut	pangkuk	bengkut	bungkut	mengkut	pengkut	bengkut
	DC6	pakan	bekan	-	-	-	-	-
	DC7	bangkun	gantung	panji	bangkang	bangkan	bankan	kantung
	DC8	berkur	bangkur	bengkul	bengkun	bekul	belkol	bekung
	DC9	benker	pergi	-	-	-	-	-
	DC10	benikuan	dangikuan	blengkuan	-	-	-	-

Target		**Session 1**	**Session 2**	**Session 3**	**Session 4**	**Session 5**	**Session 6**	**Session 7**
tengking	DC1	tangkai	pangjing	tangkin	cankit	tangkin	tangkin	tangking
	DC2	tengkit	tenkit	tekit	tengkai	tengkit	tengkit	tengkit
	DC3	tekang	tengking	tengking	tengjiring	terjidi	tengkat	tengking
	DC4	tingkah	tangki	tingka	tekak	tingki	tekik	tengki
	DC5	tengkit	tengkit	tengkit	tengkit	tengkit	tengkit	tengkit
	DC6	takan	tentik	-	-	-	-	-
	DC7	te'ngji	te'nko	tangkut	tangkun	tangkur	tantun	tantong
	DC8	terte'k	terting	tengkai	tingkap	teng_ing	tempat	tengking
	DC9	te'nkan	te'rin	-	-	-	-	-
	DC10	tenakian	tengkian	tengkian	-	-	-	-

Syllable Pattern: PAIRED VOWELS

Target		**Session 1**	**Session 2**	**Session 3**	**Session 4**	**Session 5**	**Session 6**	**Session 7**
saut	DC1	suat	suat	suat	suat	suat	suat	suat
	DC2	satu	satu	satu	satu	sabtu	sabtu	satu
	DC3	sentut	sutu	satu	saut	satut	sup	sut
	DC4	sut	sut	sut	sut	sut	sut	susat
	DC5	saut	sua	suat	sanggut	sa_ut	sa_ut	sa_ut
	DC6	sap	senuk	-	-	-	-	-
	DC7	sut	kut	sut	sut	kasut	kasut	kasut
	DC8	satu	satu	satu	satu	satu	satu	satu
	DC9	situ	satu	-	-	-	-	-
	DC10	satu	satu	satu	-	-	-	-

Target		**Session 1**	**Session 2**	**Session 3**	**Session 4**	**Session 5**	**Session 6**	**Session 7**

puas	DC1	puas	pus	pus	pus	pus	puas	puas
	DC2	pusa	unsa	bunsat	punsat	pusat	pusat	pusat
	DC3	pas	puas	puas	upas	puas	puas	puas
	DC4	paus	puas	paus	puas	puas	puas	puas
	DC5	pu_as	pu_as	pu_as	puas	bu_as	puas	puas
	DC6	pu_as	pusa	-	-	-	-	-
	DC7	pas	pus	pas	bas	bas	pos	sakop
	DC8	pua	puas	pesas	pasa	buas	puas	puhas
	DC9	pada	pasu	-	-	-	-	-
	DC10	pusa	puas	pusa	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
suap	DC1	suap	sup	suap	suap	suap	suap	suap
	DC2	supan	tumpah	sup	patuk	suda	supa	puap
	DC3	sepu	sup	-	-	sup	sup	sut
	DC4	sup	sup	sup	sup	sup	sup	sup
	DC5	su_ap	su_ap	su_ap	su_ap	suap	suap	sa_up
	DC6	sapap	supa	-	-	-	-	-
	DC7	paps	puts	sap	sarus	sap	sup	suya
	DC8	pas	supas	pua	upa	puas	suap	supa
	DC9	supar	sapu	-	-	-	-	-
	DC10	sapu	sapu	supa	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
buas	DC1	buas	pus	buas	bus	wus	buas	puas
	DC2	busa	bunsa	sumpah	bansai	punsa	batuk	pu
	DC3	bus	bus	buas	udas	spasu	udas	bus
	DC4	bas	bask	bas	bas	bas	bas	bas
	DC5	pu_as	bu_as	pu_as	bu_as	puas	buas	paus
	DC6	tas	bu_as	-	-	-	-	-
	DC7	bas	puats	bas	pas	bas	bas	bas
	DC8	buhat	buas	buas	buah	buas	buap	buah
	DC9	ban	te'rin	-	-	-	-	-
	DC10	tenakian	tengkian	tengkian	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
paut	DC1	puat	put	put	pau	puat	puat	put
	DC2	bantu	patuk	putsa	patuk	batu	pusat	pa
	DC3	pun	pus	pun	pah	pintu	piut	puk
	DC4	patut	paut	puta	patut	patut	patut	patut
	DC5	pau	pa_ut	pa_ut	puat	puat	pa_ut	paut
	DC6	put	pu_ut	-	-	-	-	-
	DC7	pur	put	por	put	put	bit	pat
	DC8	puta	patut	puta	putoh	pusat	pusat	utop
	DC9	put	patu	-	-	-	-	-
	DC10	patu	puas	put	-	-	-	-

Syllable Pattern: CV+CV+CV

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
selesa	DC1	selasa	selasa	selasa	selasa	selasa	selasa	selasa
	DC2	selisa	selisa	selisa	selasa	selasa	selisa	selasa
	DC3	selasa	selasa	selasa	selasa	selasa	selasa	selasa
	DC4	sele'sa	sele'sa	sele'sa	sele'sa	sele'sa	sele'sa	sele'sa
	DC5	terasa	selasa	selasa	selasa	selasa	selasa	selasa
	DC6	salsa	temasa	-	-	-	-	-
	DC7	selag	seldik	selit	selesi	seling	selisa	silang
	DC8	selesai	selesai	selesai	selesai	selesai	selesai	selesai
	DC9	selase	selisa	-	-	-	-	-
	DC10	selasa	selasa	selasa	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
ketara	DC1	ketara	ketara	ketara	ketara	ketara	ketara	ketara
	DC2	ketara	kitara	ketara	ketara	ketara	ketara	ketara
	DC3	ketar	ketara	terka	kereta	-	kereta	kere'ta
	DC4	-	kerata	keratu	ketare	ketara	tara	ketarang
	DC5	ketara	ketara	ketara	ketara	kedara	ketara	ketara
	DC6	-	kera	-	-	-	-	-
	DC7	ketar	ketir	ketar	ketari	ketir	ketira	ketar
	DC8	ketira	ketara	ketara	kenata	ketada	ketara	ketara
	DC9	akan	ketas	-	-	-	-	-
	DC10	ketara	ketara	ketara	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
menara	DC1	mara	menara	menara	menara	menara	menara	menara
	DC2	menara	menara	menara	menara	menara	menara	menara
	DC3	merana	merana	merana	merana	merana	merana	merana
	DC4	bende'ra	menari	menari	menari	menari	menari	menari
	DC5	melara	menara	menara	menara	menara	menara	menara
	DC6	mera	merara	-	-	-	-	-
	DC7	benra	banar	menar	menari	menari	menari	menar
	DC8	menara	menarang	menarai	menari	menara	menara	menara
	DC9	menyara	merari	-	-	-	-	-
	DC10	menara	menara	menara	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
kepala	DC1	kepala	kepala	kepala	kepala	kelapa	kepala	kepala
	DC2	kepala	kepala	kepala	kelapa	kepala	kepala	kepala
	DC3	kepala	kelapa	kepala	kelapa	pala	kepala	kelapa
	DC4	kepale	kepala	kelape	kepala	kepale	kepale	kepala
	DC5	kepada	kepala	kepala	kepala	kepala	kepala	kepala
	DC6	ke'la	kepala	-	-	-	-	-
	DC7	kepal	kepala	kepal	kepala	kepal	kepala	kepala
	DC8	kepata	kepala	kepala	kepala	kepala	kepala	kepala
	DC9	kila	kipas	-	-	-	-	-
	DC10	kepala	kepala	kepala	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
kelapa	DC1	kelapa	kelapa	kelapa	kelapa	kelapa	kelapa	kelapa
	DC2	kelapa	kelapa	kelapa	kepala	kelapa	kelapa	kelapa
	DC3	kelapa	kepala	kelapa	kepala	kelapa	kelapa	kepala
	DC4	lapa	kelapa	kelape	kelapa	kelape	kelapa	kelapa
	DC5	kelapan	kelapa	kelapa	kelapa	kelapa	kelapa	kepala
	DC6	apa	kapa	-	-	-	-	-
	DC7	kelap	kelap	kepal	kelapa	kelapa	kelapa	kelap
	DC8	kelupa	kelapa	kelapa	kelapa	kelapa	kelapa	kelapa
	DC9	kelapar	kelto	-	-	-	-	-
	DC10	kelapa	kelapa	kelapa	-	-	-	-

Syllable Pattern: CV+CV+CCV

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
telinga	DC1	telinga	telinga	telinga	telinga	telinga	telinga	telinga
	DC2	telinga	telinga	telinga	telinga	telinga	telinga	telinga
	DC3	tinggal	tinggal	tinggi	tetinggal	tetinggal	tinggi	telinga
	DC4	telina	terina	belina	telinga	telina	telinga	telina
	DC5	telingak	telingga	teliga	telingga	telingga	tetingga	tetingna
	DC6	telinga	tengaja	-	-	-	-	-
	DC7	-	telag	telag	telna	telingga	lingga	telang
	DC8	telin_gar	telinggal	telingai	telinga	telinga	telinga	telin_gai

	DC9	tengang	tal_ing	-	-	-	-	-
	DC10	telingga	telinggi	telingai	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
pelangi	DC1	pelangi	pelangi	pelangi	pelangi	pelangi	pelangi	pelangi
	DC2	belanang	pelalang	pelanang	belalang	pelangki	pelangki	pelakin
	DC3	palang	pelang	perlang	melanggar	membelai	pewani	pilang
	DC4	pelani	pelanik	pelanit	pelanit	pelanik	pelani	pelanik
	DC5	pelangkit	pelanggit	pelanggi	pelanggi	pelangkit	pelangkit	pelangkit
	DC6	pangi	pelani	-	-	-	-	-
	DC7	pelag	pelag	pelag	perag	pelang	pelingga	pele'ng
	DC8	peganjar	pedeli	penanggi	pelang	penagai	pelagai	pelagai
	DC9	pagi	padi	-	-	-	-	-
	DC10	pilanga	pelan_gi	peragi	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
belanga	DC1	belanga	pelinga	pelangi	balang	bilanga	belanga	belanga
	DC2	belana	belalang	belalang	belakat	belagak	belangka	pelan_ga
	DC3	malang	bilang	bilang	mengilang	bilang	belalang	belang
	DC4	melangi	belalang	pelanggi	banglangga	belangga	langga	beraling
	DC5	pelangkat	pelangga	pelangkat	belangga	belangga	belangka	pelangka
	DC6	be'linga	belanga	-	-	-	-	-
	DC7	belag	belang	pelang	beEDa	belanggi	pelingga	bele'ng
	DC8	beluja	belangja	belalang	melangga	belalang	belalai	belagai
	DC9	berdangah	belnaga	-	-	-	-	-
	DC10	belangnga	belangga	belaga	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
berenga	DC1	rena	berangai	peranga	peranga	biranga	beranga	beranga
	DC2	perenang	berina	berekat	terenak	berenga	ber engka	berenang
	DC3	berenang	beranga	bera	merana	cira	merana	perang
	DC4	berang	meranggang	berangga	biring	berangga	berangga	berangga
	DC5	berekat	belrengga	berengka	berengga	berengga	berengkak	perengkak
	DC6	panga	beranga	-	-	-	-	-
	DC7	berangk	bering	perang	berega	berangga	bere'ngga	bere'ng
	DC8	beragai	berangnang	berengai	beranggai	berenjang	berigai	meregai
	DC9	berangan	berniaga	-	-	-	-	-
	DC10	berangga	berangga	berengga	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
kenanga	DC1	-	kenanga	-	kenanja	kenanga	kenanga	kenanga
	DC2	ke nan ga	ke nan ga	kenarnga	kenakat	kenan ga	kenang ga	kenanang
	DC3	mena	mera	perang	kepanan	kenan	kempunan	terana
	DC4	kenaan	kenaan	kuna	kenaan	ke_anan	kenang	kenaan
	DC5	kenangka	kenangga	kenangga	kenagga	kenagga	kenagga	kenangga
	DC6	kenga	kelanga	-	-	-	-	-
	DC7	kan_gan	kennang	kenang	kenagi	kenaggi	kenangja	kene'ngka
	DC8	kenajar	kenangjar	kenanai	kenaggai	kenajang	kenagai	kenagai
	DC9	perkalkan	kennyang	-	-	-	-	-
	DC10	kenaga	kenanara	kenaga	-	-	-	-

Syllable Pattern: CVC+CV+CVCC

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
pendatang	DC1	pendatan	pemandatan	pendatang	pendatang	pendatang	pendatang	pendatang
	DC2	mendatang	pendata	mendanang	pendapat	pendadak	pendatar	pendatang
	DC3	pendatang	pendatang	bintang	pendatang	pang	pendatang	petang
	DC4	pedanatan	datanan	pendatang	pendatang	pendatang	pendatang	pendatangan
	DC5	pendatak	pendatang	pedangtah	pendatang	pendatang	pendatang	pendatang
	DC6	panga	pematan	-	-	-	-	-

	DC7	pantag	pantang	pendatang	pendatang	pendatang	pendatang	tandata
	DC8	dedate'	pedatang	pedatai	pendapai	pendapat	pendatai	senbate'k
	DC9	pen'ntars	pe'nda	-	-	-	-	-
	DC10	pedatangan	pendatangan	pendatang	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
tembelang	DC1	tembalang	tembalang	tembelang	tembalang	tebalang	tembalang	tembalang
	DC2	tembilang	tembilak	tembilang	tembila	tembilak	tembelang	tembelang
	DC3	memhilang	terbilang	terladang	terlambat	tela	terbilang	terbilang
	DC4	tiba	telaba	petali	tebelang	tebelang	tembilang	tembalang
	DC5	tembelak	tembe'lang	tembelak	tebelang	tebelang	tembelang	tembelang
	DC6	belang	tebarang	-	-	-	-	-
	DC7	penbalik	tampun	tempanla	tembalik	tambor	tembor	tembele'ng
	DC8	tembela	tembalang	kebelang	tembelah	tembelang	tembelang	tembelang
	DC9	tembolan	te'ntatnya	-	-	-	-	-
	DC10	tembelangan	membeliang	tembelang	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
temberang	DC1	tembarang	tembirang	temberang	tembarang	terbaring	terbaring	temparang
	DC2	tember an	tembirak	timbirang	tembira	tembir ang	temberang	temperang
	DC3	merana	termelang	terdengar	tembara	-	terbawa	terbarang
	DC4	tebara	terbara	cepali	tembarang	te'rbara	tembarang	tembarang
	DC5	temberam	tembe'rang	tembere'	temberang	teberang	tembe'rang	temberang
	DC6	barang	tebarang	-	-	-	-	-
	DC7	penbarag	tampur	tembari	temburga	tambur	tembere'ng	tambor
	DC8	tembere'ng	teberang	temrangai	temkejar	temberi	tembarang	tembering
	DC9	tembaran	te_mbikai	-	-	-	-	-
	DC10	temberangan	tember_ang	temberangan	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
tempurung	DC1	-	terpurung	terpunung	terpurung	tempurung	terpurung	terpurung
	DC2	tempuyung	tempuruk	temburung	tempurai	temburuk	tempurung	tempurung
	DC3	mengurung	tempurang	terpurung	temkuran	terubu	-	terowong
	DC4	tepuru	terauha	tambar	tempurung	uru	tepuru	temparung
	DC5	telpurung	tempurung	teburung	tempurung	temurung	tempurung	tempurung
	DC6	parung	tebarang	-	-	-	-	-
	DC7	panpurnag	tembur	tempur	tempur	tampur	tempurung	merong
	DC8	temperung	teperung	tempurung	tempurung	tempurun	tempurang	tempirung
	DC9	tempuleran	komputer	-	-	-	-	-
	DC10	tempurangan	tempurungan	tempuruan	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
pembayang	DC1	pembanyak	pembayang	pembayang	pembayang	pembayang	pembayang	pembayang
	DC2	pembayang	pembayang	pembawa	pembawa	pimbayak	pembawa	pemayang
	DC3	membayang	pembayang	terbayang	tumbayang	tere'wang	pembayar	pembayang
	DC4	pebayang	pebayang	pembayang	pembayang	pembayang	pembayang	pembayang
	DC5	pembayah	penbayang	pembayang	pembayang	pembayang	pembayang	pembayang
	DC6	parang	berbasan	-	-	-	-	-
	DC7	tumbayag	pembai	pembaya	pembakan	pambaya	pemboya	panbaya
	DC8	pembagai	tembawang	pembawang	pengbawai	pembanya	pembawai	pembawang
	DC9	bernyapayan	perbanyak_an	-	-	-	-	-
	DC10	pemdayangan	pembayangan	pembayang	-	-	-	-

Syllable Pattern: CVCC+CV+CVC

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
bungkusan	DC1	bungkusan	bungsan	bungkas	bungkus	pembungkus	bungkusan	bungkusan
	DC2	pungkus	bunkusan	bunungkusan	bungkus	bungkusan	bungsat	pungkusan
	DC3	mengkusang	terbayang	berkusang	memkusan	bupusang	-	bungkusan
	DC4	bunggusan	bungkusan	kusan	bukusan	bukusan	bungkusan	bukusan

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
	DC5	bungkus	bungkussan	bukussan	mungkusan	bungkusan	bungkusen	bungkusan
	DC6	berkus_an	pebarang	-	-	-	-	-
	DC7	bangkus	gantung	buts	basan	bongkus	bangkus	pangkor
	DC8	berkuse'r	bangkutsan	bangkasan	bukusan	bengkasan	bangkusan	bukusan
	DC9	bukanya	bersukan	-	-	-	-	-
	DC10	dungankusan	bugakusan	bungkisan	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
bangsawan	DC1	pensawan	bangsawan	pensawang	bangsawan	bangsawan	bangsawan	bangsawang
	DC2	bangsa	bansawi	bansawi	bansau	bangsawi	bangsawin	pengsawin
	DC3	sawah	tan	dansawan	memsawah	sawang	dansawah	bangsawan
	DC4	basi	bangsawah	bangsawi	bangsui	bangsui	bangsawan	bangsawan
	DC5	bangsawan	bangsawah	bangsasan	bangsawan	bangsawan	bangsawan	bangsawan
	DC6	basawan	bebawan	-	-	-	-	-
	DC7	bengksawa	bangsai	nganwas	bangwan	bangsawang	bongsawang	bangkowan
	DC8	bawase'r	besawi	bengsanai	bengsawang	bengsawang	bengsawang	basawak
	DC9	bungsu	besawa	-	-	-	-	-
	DC10	bagawasan	bagawasan	bungsawan	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
jangkitan	DC1	jangkitan	janggitan	jangkitan	janggitan	jangitan	jenitan	jangkitan
	DC2	jangkitin	jangkitdan	jankitan	jankit	jangkitan	jangkittan	jangkitan
	DC3	jekatan	-	jingkitan	jangkitan	jangkitan	jangkitan	jangkitan
	DC4	jangkit	jangkit	jangkit	jangkit	jangkit	jangkit	jangkit
	DC5	jangkittan	jangkitan	jangkitan	jangkitan	jangkitan	jangkitan	jangkitan
	DC6	jambatan	jakatan	-	-	-	-	-
	DC7	jangkit	jaya	bankusa	jangkitang	jangkina	jangtang	jenkot
	DC8	jengtite'ng	jemketa	jemtika	pendika	jemtikan	jemtikkan	jentika
	DC9	jukan	jatan	-	-	-	-	-
	DC10	jangkitan	jangkitan	jangkitan	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
bingkisan	DC1	pengkasan	pangkisan	tangisan	mengkisan	pangkisan	bangkisan	mengkisan
	DC2	binkisan	bingkisan	bankisan	binkisan	bingkisan	bengkisan	bingkisan
	DC3	mengkisang	mengkasian	bangkisan	bisankan	bekisar	-	bingkisan
	DC4	balkis	bakis	bisik	pangsi	bangsi	bangsik	pakis
	DC5	bingkisan	bingkitsan	bingkisan	bingkikan	bingkisan	bingkisan	bingkisan
	DC6	basan	berisan	-	-	-	-	-
	DC7	bingkisai	bangsan	bangkuts	bingkinsan	bangkuns	banting	pantok
	DC8	bengkise'r	bengkisang	bingkisan	bingkisan	bingkisan	bengkisan	bingkisan
	DC9	berkas	bijaks	-	-	-	-	-
	DC10	dagikisan	bikisan	binkisan	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
panggilan	DC1	penjali	panggilan	panggilan	pangngilan	panggilan	panggilan	panggilan
	DC2	panggil	panggil	panggian	panggil	panggilan	panggilan	panggilan
	DC3	menghilang	porkapail	panggilan	memninggal	meninggal	-	pemlalang
	DC4	panggilan	panggilan	panggilan	pe'nggilan	panggilan	panggilan	panggilan
	DC5	panggilan	bangkilan	bangkilan	panggislan	mangkilan	bangkilan	panggilan
	DC6	palan	pegilan	-	-	-	-	-
	DC7	bangki	panglun	panje'	pangile'ng	pangling	paling	panglong
	DC8	penggela	pengjalang	panggilan	pengdilai	pengelang	panggilan	panggilan
	DC9	peragalan	panggang	-	-	-	-	-
	DC10	pangie'ian	panggilan	Pangkasan	-	-	-	-

Syllable Pattern: CV+CVCC+CVC

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
perangkap	DC1	perangkap	perangkap	perangkap	perangkap	perangkap	perangkap	perangkap
	DC2	per an kap	berangkap	merangkat	perakap	ber angkap	per angkap	ber_angkap

	DC3	cakap	-	bercakap	bercakap	bercakap	bercakap	bercakap
	DC4	perangkap	perangkap	perangkap	perangkap	perangkap	perangkap	prangkap
	DC5	perangkatkan	perangkat	perangkat	perangkap	perangkat	perangkap	perangkap
	DC6	takap	perikap	-	-	-	-	-
	DC7	bekap	perag	peragkur	perangkap	perangkam	pere'ngkam	peringkap
	DC8	perakut	perangtang	perantap	perakap	perangkap	perakai	perakan
	DC9	percapancan	perjagar	-	-	-	-	-
	DC10	per_ang_kap	peranggikap	Perangkat	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
merengkok	DC1	merokoi	meringkok	merongkok	perongkok	perongkok	merengkok	meringkok
	DC2	merinengkok	menrikuk	mereko	merenangkok	maringkok	mereko	merekok
	DC3	merangkak	merokok	berkongkong	merokok	merokok	memrokok	merokok
	DC4	meraga	merokok	merongga	merokok	merongkong	merongga	murokok
	DC5	merengpak	me're'ngkut	merengkok	merengkok	merengkok	merengkok	merongkok
	DC6	mongkok	merokok	-	-	-	-	-
	DC7	meragko	morkot	meragko	merangko	merangko	merengko	pere'ngko
	DC8	merodok	merongkok	merakur	merangkor	merengkor	merakur	merakor
	DC9	merdarkot	merjagor	-	-	-	-	-
	DC10	merangkikak	mereke'kok	merangkok	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
selongkar	DC1	selongkar	serikon	selongkok	serangkai	selongkar	selongkok	selongkar
	DC2	selolongkar	selopak	selongkar	selonoerkar	selokar	selokar	selongkar
	DC3	sukar	sokoran	sengkolar	selongkar	salar	sengkoran	selongkar
	DC4	sorongga	selonggar	selonggar	solongkar	selanggor	longgar	selonggar
	DC5	selongkar	selongkat	selongko	selongkar	selongkar	selongka	selowangkak
	DC6	sakoga_ar	sekokar	-	-	-	-	-
	DC7	selogra	selog	selogo	selangkor	selangkor	selangkor	selongkor
	DC8	selajar	selangkor	selangkai	selakar	selesai	selorong	selorai
	DC9	salokan	sellokar	-	-	-	-	-
	DC10	selogokar	selonggokar	selongkar	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
melanggar	DC1	melanga	melangan	melangar	merangar	merangar	melangar	melangar
	DC2	melanggar	melanggar	belanggar	melangar	melanak	langgar	melanggar
	DC3	melayar	menyinar	meninar	meninggal	meral	meninggal	langgar
	DC4	melanggar	melanggar	melanggar	melanggar	melanggar	melanggar	melanggar
	DC5	melanggar	melanggar	melanggar	melanggar	melamnggar	melanggar	melangga
	DC6	mengga	mere'gar	-	-	-	-	-
	DC7	melag_rag	malang	melag	melanggi	melangging	mele'ngkor	melinggan
	DC8	melangjar	melanggai	melanggai	melanggar	melanggai	melarai	melagai
	DC9	melalohir	melagar	-	-	-	-	-
	DC10	melanggaran	melanggar	melanggar	-	-	-	-

Target		Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
merangkak	DC1	merangkak	merangkak	merangkak	merangkak	meringkak	merangkak	merangkak
	DC2	mer an ka	merankal	merangkap	meranangkak	merangka	merangkat	merangkat
	DC3	merangkak	merangkak	merangkak	merangkak	merangkak	merangkak	merangkak
	DC4	merakak	merakak	merakak	merakak	merangkak	merakak	merakak
	DC5	merangkat	merangkak	merangka	merangkak	merangga	merakap	merakkak
	DC6	mere'kak	mere'kak	-	-	-	-	-
	DC7	merangk	merankit	meragko	merangkan	menangkun	mere'ngkor	mere'ngkan
	DC8	merukit	merangkak	merangkak	merangkak	merangkak	merangkak	merangkak
	DC9	merikat	bernyakak	-	-	-	-	-
	DC10	mer_an_ka	merangkak	merangkak	-	-	-	-

Appendix C

Percentage of Occurrences of Modeled Pronunciations

Target	Transcript	N=56	Frequency (%)
aku	ku	51	91.07
	ape	5	8.93
	apa	1	1.79
	aku	1	1.79
	aka	1	1.79

Target	Transcript	N=56	Frequency (%)
apa	bace	49	87.50
	aki	7	12.50

Target	Transcript	N=56	Frequency (%)
baca	baca	47	83.93
	apa	2	3.57
	baci	1	1.79
	baka	1	1.79
	bica	1	1.79
	baja	1	1.79
	beca	1	1.79
	paca	1	1.79
	bapak	1	1.79

Target	Transcript	N=56	Frequency (%)
suka	suka	53	94.64
	nyata	1	1.79
	ayat	1	1.79
	hayati	1	1.79

Target	Transcript	N=56	Frequency (%)
nyata	nyati	35	66.04
	menata	2	3.77
	entaya	2	3.77
	ya	1	1.89
	nata	1	1.89
	naya	1	1.89
	ayatan	1	1.89
	anatan	1	1.89
	nukmat	1	1.89
	nagta	1	1.89
	anya	1	1.89
	nyatah	1	1.89

	anyata	1	1.89
	tanya	1	1.89
	duka	1	1.89
	sukar	1	1.89
	suke	1	1.89

Target	Transcript	N=56	Frequency (%)
bunga	bunga	26	46.43
	bujang	5	8.93
	buna	4	7.14
	buji	4	7.14
	abang	2	3.57
	bun_ga	1	1.79
	buga	1	1.79
	bunja	1	1.79
	bunge	1	1.79
	tanga	1	1.79
	bena	1	1.79
	banyak	1	1.79
	bangga	1	1.79
	punka	1	1.79
	benja	1	1.79
	unah	1	1.79
	bugan	1	1.79
	boja	1	1.79
	banda	1	1.79
	puhkan	1	1.79

Target	Transcript	N=56	Frequency (%)
ceria	ceria	36	64.39
	curia	3	5.36
	cerai	2	3.57
	kiri	2	3.57
	ceda	2	3.57
	cira	1	1.79
	cari	1	1.79
	cera	1	1.79
	kerna	1	1.79
	cerah	1	1.79
	caria	1	1.79
	cerin	1	1.79
	cili	1	1.79
	ceri	1	1.79
	ringgi	1	1.79

Target	Transcript	N=56	Frequency (%)
wad	wad	20	36.36

	wan	13	23.64
	wap	11	20.00
	yad	2	3.64
	we‘p	1	1.82
	ding	1	1.82
	wa	1	1.82
	wam	1	1.82
	man	1	1.82
	we‘n	1	1.82
	wab	1	1.82
	kowap	1	1.82
	wed	1	1.82

Target	**Transcript**	**N=56**	**Frequency (%)**
umur	umur	32	60.38
	amur	4	7.55
	mur	3	5.66
	umir	3	5.66
	yuruh	1	1.89
	umura	1	1.89
	umo	1	1.89
	abur	1	1.89
	mukar	1	1.89
	pame‘r	1	1.89
	yuror	1	1.89
	amar	1	1.89
	upur	1	1.89
	mmur	1	1.89
	umer	1	1.89

Target	**Transcript**	**N=56**	**Frequency (%)**
abah	abah	28	50.00
	adah	6	10.71
	abu	5	8.93
	ada	4	7.14
	abak	2	3.57
	ubah	2	3.57
	apa	1	1.79
	aha	1	1.79
	abih	1	1.79
	baba	1	1.79
	abur	1	1.79
	abor	1	1.79
	apu	1	1.79
	abang	1	1.79

Target	**Transcript**	**N=56**	**Frequency (%)**
ayat	ayat	17	30.91

	ayah	16	29.09
	nyata	4	7.27
	ayi	4	7.27
	tanya	3	5.45
	ayap	2	3.64
	ayam	2	3.64
	ya	1	1.82
	tayat	1	1.82
	anyat	1	1.82
	yang	1	1.82
	ngeti	1	1.82
	atah	1	1.82
	anya	1	1.82

Target	**Transcript**	**N=56**	**Frequency (%)**
kemarau	kemaru	9	16.07
	kemarau	7	12.50
	kemarai	7	12.50
	kemariau	4	7.14
	kemar	3	5.36
	kemurang	2	3.57
	kemurangan	2	3.57
	kemeliau	1	1.79
	kemurahan	1	1.79
	kamar	1	1.79
	kemaran	1	1.79
	kemuru	1	1.79
	kemuruh	1	1.79
	kemurun	1	1.79
	kemarou	1	1.79
	kerok	1	1.79
	kemeran	1	1.79
	umur	1	1.79
	kemarur	1	1.79
	kemaruai	1	1.79
	kemuraran	1	1.79
	kemuruh	1	1.79
	kemmaru	1	1.79
	kemerai	1	1.79
	pemarai	1	1.79
	pemmaru	1	1.79
	kemurai	1	1.79
	kemmari	1	1.79
	kemerai	1	1.79

Target	**Transcript**	**N=56**	**Frequency (%)**
kumpulan	kumpulan	23	41.07

	kempulan	4	7.14
	kupulan	4	7.14
	kumpulang	3	5.36
	kumpulong	1	1.79
	kamponglan	1	1.79
	kempun	1	1.79
	tempatlah	1	1.79
	kempul	1	1.79
	kepulau	1	1.79
	kembulong	1	1.79
	kemula	1	1.79
	kempur	1	1.79
	punglan	1	1.79
	kempola	1	1.79
	kemputtan	1	1.79
	jemputlan	1	1.79
	kepulangan	1	1.79
	lupaan	1	1.79
	kampunan	1	1.79
	kempulen	1	1.79
	pulangan	1	1.79
	kepuloi	1	1.79
	kumpunan	1	1.79
	kumbulan	1	1.79
	kempulang	1	1.79

Target	**Transcript**	**N=56**	**Frequency (%)**
cendawan	cendawan	32	57.14
	cedawan	7	12.50
	cendan	1	1.79
	canwan	1	1.79
	sedawan	1	1.79
	cadwan	1	1.79
	hantu	1	1.79
	cedawa	1	1.79
	sindiawan	1	1.79
	dedawan	1	1.79
	sandawa	1	1.79
	cedawai	1	1.79
	cindawan	1	1.79
	sendawan	1	1.79
	kandawan	1	1.79
	sandawan	1	1.79
	jembawang	1	1.79
	jengdawan	1	1.79
	cawangan	1	1.79

Target	**Transcript**	**N=55**	**Frequency (%)**

maklumat	maklumat	24	43.64
	makulmat	4	7.27
	malumat	3	5.45
	kelumat	3	5.45
	mak_umat	2	3.64
	maklong	2	3.64
	muluat	1	1.82
	mamut	1	1.82
	melumat	1	1.82
	perlautan	1	1.82
	meluan	1	1.82
	mengkeliatan	1	1.82
	makkumat	1	1.82
	aknumat	1	1.82
	manglong	1	1.82
	kelubak	1	1.82
	mak l mat	1	1.82
	makelmat	1	1.82
	melut	1	1.82
	kalum	1	1.82
	makwan	1	1.82
	kelua	1	1.82
	kantum	1	1.82

Target	Transcript	N=55	Frequency (%)
belalang	belalang	23	41.82
	balang	4	7.27
	belang	4	7.27
	bilang	3	5.45
	belakang	2	3.64
	belalai	2	3.64
	pelalang	2	3.64
	belanglang	2	3.64
	pulan	1	1.82
	belagak	1	1.82
	berlawan	1	1.82
	lalang	1	1.82
	berula	1	1.82
	beedi	1	1.82
	belaling	1	1.82
	belag	1	1.82
	belale'ng	1	1.82
	belangkan	1	1.82
	berlanang	1	1.82
	pelangkan	1	1.82
	belangka	1	1.82

Target	Transcript	N=56	Frequency (%)

kecundang	kecundang	14	25.00
	kecundan	3	5.36
	kecedang	2	3.57
	kecudang	2	3.57
	kecudan	1	1.79
	belatan	1	1.79
	kecundangu	1	1.79
	kicapdan	1	1.79
	kecudangan	1	1.79
	kecondok	1	1.79
	kenada	1	1.79
	kucundang	1	1.79
	kenacang	1	1.79
	kecunda	1	1.79
	kecundai	1	1.79
	e‘cunanan	1	1.79
	kekunding	1	1.79
	kecunduang	1	1.79
	kucawan	1	1.79
	kecindung	1	1.79
	kecungdang	1	1.79
	kuncang	1	1.79
	kecundangan	1	1.79
	kecendang	1	1.79
	pencudang	1	1.79
	kecandang	1	1.79
	kecunju	1	1.79
	kejudang	1	1.79
	kelongda	1	1.79
	kecuadan	1	1.79
	sedang	1	1.79
	kesunnga	1	1.79
	kucitikan	1	1.79
	kecundah	1	1.79
	kedang	1	1.79
	sunda	1	1.79
	kecunjingkan	1	1.79
	kelongda	1	1.79
	kekanggi	1	1.79

Target	Transcript	N=56	Frequency (%)
orang	orang	47	83.93
	oren	4	7.14
	ore‘n	3	5.36
	urang	1	1.79
	upah	1	1.79

Target	Transcript	N=56	Frequency (%)

abang	abang	36	64.29
	adang	4	7.14
	atuk	1	1.79
	benggugeng	1	1.79
	angan	1	1.79
	adangu	1	1.79
	amang	1	1.79
	bang	1	1.79
	abangan	1	1.79
	ubangnga	1	1.79
	abing	1	1.79
	abong	1	1.79
	adangan	1	1.79
	api	1	1.79
	abo	1	1.79
	abah	1	1.79
	apa	1	1.79
	abu	1	1.79

Target	Transcript	N=55	Frequency (%)
bawang	bawang	37	67.27
	bawa	5	9.09
	bewang	2	3.64
	badan	1	1.82
	payang	1	1.82
	berlaku	1	1.82
	bewangan	1	1.82
	bewa	1	1.82
	bawangan	1	1.82
	bawat	1	1.82
	barang	1	1.82
	bayang	1	1.82
	bawah	1	1.82
	awang	1	1.82

Target	Transcript	N=56	Frequency (%)
jangan	jangan	18	32.14
	janan	7	12.50
	jangkat	6	10.71
	jangat	2	3.57
	jang an	1	1.79
	ajang	1	1.79
	jengan	1	1.79
	jakin	1	1.79
	jakit	1	1.79
	jenang	1	1.79
	jangkan	1	1.79
	jengja	1	1.79

	janggan	1	1.79
	jajan	1	1.79
	juga	1	1.79
	jantan	1	1.79
	jangga	1	1.79
	janjan	1	1.79
	jalan	1	1.79
	jangkan	1	1.79
	jemputan	1	1.79
	jantak	1	1.79
	jengjan	1	1.79
	jemput	1	1.79
	jenja	1	1.79

Target	Transcript	N=55	Frequency (%)
sayang	sayang	36	65.45
	saya	5	9.09
	sayur	4	7.27
	sanya	2	3.64
	senge	1	1.82
	sayap	1	1.82
	layang	1	1.82
	sawan	1	1.82
	sarang	1	1.82
	seyi	1	1.82
	sungai	1	1.82
	cayang	1	1.82

Target	Transcript	N=56	Frequency (%)
barang	barang	38	67.86
	barangan	3	5.36
	baring	3	5.36
	bare'ng	2	3.57
	badan	1	1.79
	berna	1	1.79
	bewat	1	1.79
	berkan	1	1.79
	birang	1	1.79
	barkar	1	1.79
	berang	1	1.79
	parang	1	1.79
	burung	1	1.79
	bawang	1	1.79

Target	Transcript	N=56	Frequency (%)
betul	betul	37	66.07
	botol	5	8.93
	belur	2	3.57

	betal	2	3.57
	belum	1	1.79
	betuk	1	1.79
	beli	1	1.79
	butul	1	1.79
	belu	1	1.79
	buzutan	1	1.79
	batak	1	1.79
	betel	1	1.79
	betol	1	1.79
	beltul	1	1.79

Target	**Transcript**	**N=55**	**Frequency (%)**
pandai	pandai	20	36.36
	pantai	5	9.09
	pandi	4	7.27
	pandan	3	5.45
	pandang	2	3.64
	banding	2	3.64
	pedang	2	3.64
	bantai	2	3.64
	pebagai	1	1.82
	pendang	1	1.82
	pelan	1	1.82
	pedi	1	1.82
	padah	1	1.82
	pidai	1	1.82
	pedal	1	1.82
	andai	1	1.82
	panti	1	1.82
	penggan	1	1.82
	padanya	1	1.82
	pendah	1	1.82
	pina	1	1.82
	pading	1	1.82
	bidai	1	1.82

Target	**Transcript**	**N=56**	**Frequency (%)**
pergi	pergi	34	60.71
	pegi	8	14.29
	bergi	3	5.36
	pargi	2	3.57
	piring	1	1.79
	peringi	1	1.79
	perag	1	1.79
	perji	1	1.79
	perga	1	1.79
	bedi	1	1.79

	perangai	1	1.79
	bera	1	1.79
	pering	1	1.79

Target	Transcript	N=56	Frequency (%)
makan	makan	51	91.07
	mekan	1	1.79
	manan	1	1.79
	makanan	1	1.79
	mamkan	1	1.79
	kan	1	1.79

Target	Transcript	N=56	Frequency (%)
pangsa	pangsa	15	26.79
	penjasa	6	10.71
	bangsa	4	7.14
	pisah	3	5.36
	pansa	2	3.57
	pasang	2	3.57
	pangsang	1	1.79
	panngas	1	1.79
	panas	1	1.79
	pangas	1	1.79
	pangsan	1	1.79
	panggas	1	1.79
	pengsu	1	1.79
	mangsa	1	1.79
	pasa	1	1.79
	angsa	1	1.79
	pakas	1	1.79
	pan_gei	1	1.79
	pingas	1	1.79
	pesa	1	1.79
	bangsan	1	1.79
	panggas	1	1.79
	pansat	1	1.79
	pisang	1	1.79
	pensa	1	1.79
	pan_gas	1	1.79
	pemasa	1	1.79
	pengsan	1	1.79
	pendahukun	1	1.79
	pengsa	1	1.79

Target	Transcript	N=56	Frequency (%)
udang	udang	38	67.86
	ubat	5	8.93
	undang	4	7.14

	ada	3	5.36
	adik	1	1.79
	lulukan	1	1.79
	ludar	1	1.79
	udah	1	1.79
	adik	1	1.79
	luga	1	1.79

Target	Transcript	N=55	Frequency (%)
jangkitan	jangkitan	18	32.73
	jangkit	8	14.55
	jemtika	2	3.64
	jatan	2	3.64
	jangkittan	2	3.64
	jengkot	1	1.82
	jentika	1	1.82
	jangtang	2	1.82
	jemtikan	1	1.82
	jangkina	1	1.82
	jangkitang	1	1.82
	pendika	1	1.82
	jingkitan	1	1.82
	bankusa	1	1.82
	jakatan	1	1.82
	jaya	1	1.82
	jemketa	1	1.82
	jekatan	1	1.82
	jambatan	1	1.82
	jengtite‘ng	1	1.82
	jukan	1	1.82
	jangkitin	1	1.82
	jangkitdan	1	1.82

Target	Transcript	N=55	Frequency (%)
pernah	pernah	17	30.36
	penah	7	12.50
	peranah	3	5.36
	perana	2	3.57
	penar	2	3.57
	enah	2	3.57
	pernat	2	3.57
	per_nah	1	1.79
	perangk	1	1.79
	pahnah	1	1.79
	perka	1	1.79
	damtai	1	1.79
	pernya	1	1.79
	pertah	1	1.79

	pernaha	1	1.79
	penang	1	1.79
	ciyernah	1	1.79
	peranai	1	1.79
	perrah	1	1.79
	pera	1	1.79
	pirham	1	1.79
	perana	1	1.79
	berah	1	1.79
	pernar	1	1.79
	perning	1	1.79
	peranak	1	1.79
	perah	1	1.79
	bernah	1	1.79

Target	Transcript	N=55	Frequency (%)
hampar	hampar	15	26.79
	hampir	10	17.86
	harapan	4	7.14
	haram	3	5.36
	hampan	2	3.57
	harmpat	2	3.57
	hampur	2	3.57
	himpir	1	1.79
	harabaq	1	1.79
	he'nper	1	1.79
	hamarkan	1	1.79
	hapar	1	1.79
	hampart	1	1.79
	hemrur	1	1.79
	parpar	1	1.79
	hambar	1	1.79
	harmpart	1	1.79
	hamra	1	1.79
	rumpar	1	1.79
	aram	1	1.79
	hempar	1	1.79
	ampar	1	1.79
	hampat	1	1.79
	hermpat	1	1.79
	hampor	1	1.79

Target	Transcript	N=56	Frequency (%)
cantik	cantik	44	78.57
	catan	1	1.79
	kante'	1	1.79
	catak	1	1.79
	natik	1	1.79

	kantik	1	1.79
	cangkit	1	1.79
	sante'ng	1	1.79
	cendik	1	1.79
	canting	1	1.79
	cindak	1	1.79
	santik	1	1.79
	nantan	1	1.79

Target	Transcript	N=56	Frequency (%)
selesa	selasa	26	46.43
	selesai	7	12.50
	sele'sa	6	10.71
	selisa	6	10.71
	sele'se	1	1.79
	terasa	1	1.79
	salsa	1	1.79
	selag	1	1.79
	selase	1	1.79
	temasa	1	1.79
	seldik	1	1.79
	selit	1	1.79
	selesi	1	1.79
	seling	1	1.79
	silang	1	1.79

Target	Transcript	N=56	Frequency (%)
kepala	kepala	38	67.86
	kepala	4	7.14
	kelape	4	7.14
	kelap	3	5.36
	kapa	2	3.57
	lapa	1	1.79
	kelapan	1	1.79
	apa	1	1.79
	kelupa	1	1.79
	kelapar	1	1.79
	kapa	1	1.79
	seldik	1	1.79

Target	Transcript	N=56	Frequency (%)
pendatang	pendatang	23	41.07
	pendatangan	3	5.36
	pendapat	2	3.57
	pen'da	1	1.79
	pedatai	1	1.79
	pendapai	1	1.79

	pendatai	1	1.79
	tandata	1	1.79
	sanbate‘k	1	1.79
	pendatan	1	1.79
	mendatang	1	1.79
	pedanatan	1	1.79
	pendatak	1	1.79
	pemandatang	1	1.79
	pendata	1	1.79
	datanan	1	1.79
	mendanang	1	1.79
	bintang	1	1.79
	pedangtah	1	1.79
	pandatang	1	1.79
	pedatang	1	1.79
	petang	1	1.79
	pendatar	1	1.79
	pendadak	1	1.79
	pang	1	1.79
	panga	1	1.79
	pantag	1	1.79
	dedate‘	1	1.79
	pe‘ntras	1	1.79
	pematan	1	1.79
	pantang	1	1.79

Appendix D

Algorithm for transcription and miscue detection in tcl

```
foreach package {Statenet Nnrun Wave TrainLibrary Garbage Feature \
      Encode Mx Viterbi Rtcl} {
      package require $package
      }

global UserComputeFeatures
global UserComposeVector


if {($argc != 10 && $argc != 12) || \
      ($argc == 10 && ![string match [lindex $argv 8] "-g"])} {
    puts "Usage: asr_edit2.tcl <.grammar> <.lexicon> <.spec> <.nnet> "
    puts "                <.wav> <.wrd> <.phn> <.cat> <target>
<filename> \[-g <garb_val>\]"
    exit
    }

set grammar_file [lindex $argv 0]
set lexicon_file [lindex $argv 1]
set spec_file    [lindex $argv 2]
set nnet_file    [lindex $argv 3]
set wav_file     [lindex $argv 4]
set wrd_file     [lindex $argv 5]
set phn_file     [lindex $argv 6]
set cat_file     [lindex $argv 7]
set target_word  [lindex $argv 8]
set filename     [lindex $argv 9]

set garbage   5
if {$argc == 10} {
      set garbage [lindex $argv 9]
      }

# read in the neural-network weights in either format
if {![catch {set fob  [obfile open $nnet_file]} msg]} {
    if {[catch {set nnet [obfile read $fob __nnet__]}]} {
        set nnet [nnet optload $nnet_file]
        }
    obfile close $fob
    } else {
    set nnet [nnet optload $nnet_file]
    }

# read in waveform
set wave [wave read $wav_file]

set stateNet [statenet create $grammar_file "grammar" \
                    -startToken {$grammar}]
statenet add $stateNet $lexicon_file "word"
statenet addSpec $stateNet $spec_file "phoneme" -selfLoops 1
statenet specUpdate $stateNet $spec_file
```

```
# statenet print $stateNet

set featuresURI [lindex [statenet info $stateNet -featuresURI] 0]
if {$featuresURI != ""} {
      regexp {(.+)#(.+)} $featuresURI whole features_file
features_proc
      source $features_file
      set UserComputeFeatures $features_proc
      }
set contextURI [lindex [statenet info $stateNet -contextURI] 0]
if {$contextURI != ""} {
      regexp {(.+)#(.+)} $contextURI whole context_file context_proc
      source $context_file
      set UserComposeVector $context_proc
      }

set sampling_freq [expr int([statenet info -sampFreq $stateNet])]
get_features $wave feat $sampling_freq
compose_feature_vector $wave $feat $sampling_freq observations

set prob  [nnet x $nnet $observations]
set probG [garbage median -N $garbage $prob]

set vObj [viterbi init $stateNet]
viterbi search $vObj $stateNet $probG
set answer [viterbi answer $vObj $stateNet -phonemes 1]
#puts $answer
viterbi reset $vObj

set msec 1.000
writeLabels $wrd_file $msec [lindex $answer 1]
writeLabels $phn_file $msec [lindex $answer 2]
writeLabels $cat_file $msec [lindex $answer 3]

puts "-------------------------------------------------------------------"
puts ""
puts $answer
puts ""
puts "Phoneme sequence recognized: "


#transcribe phoneme sequence into text
set a [list]
foreach seg [lindex $answer 2] {

      puts -nonewline "[lindex $seg 2] "

      if {[lindex $seg 2] == ".pau" | [lindex $seg 2] == "bc" |
[lindex $seg 2] == "dc" |
      [lindex $seg 2] == "gc" | [lindex $seg 2] == "kc" | [lindex $seg
2] == "pc" | [lindex $seg 2] == "tc"} {
            #do nothing
      } elseif {[lindex $seg 2] == "A"} {
            lappend a "a"
      } elseif {[lindex $seg 2] == "bh"} {
            lappend a "b"
      } elseif {[lindex $seg 2] == "tS"} {
            lappend a "c"
      } elseif {[lindex $seg 2] == "d" | [lindex $seg 2] == "dh"} {
            lappend a "d"
      } elseif {[lindex $seg 2] == "&" | [lindex $seg 2] == "E"} {
```

```
            lappend a "e"
        } elseif {[lindex $seg 2] == "f"} {
            lappend a "f"
        } elseif {[lindex $seg 2] == "g"} {
            lappend a "g"
        } elseif {[lindex $seg 2] == "h"} {
            lappend a "h"
        } elseif {[lindex $seg 2] == "I" | [lindex $seg 2] == "i:"} {
            lappend a "i"
        } elseif {[lindex $seg 2] == "dZ"} {
            lappend a "j"
        } elseif {[lindex $seg 2] == "kh" | [lindex $seg 2] == "k"} {
            lappend a "k"
        } elseif {[lindex $seg 2] == "l"} {
            lappend a "l"
        } elseif {[lindex $seg 2] == "m"} {
            lappend a "m"
        } elseif {[lindex $seg 2] == "n"} {
            lappend a "n"
        } elseif {[lindex $seg 2] == "o"} {
            lappend a "o"
        } elseif {[lindex $seg 2] == "ph"} {
            lappend a "p"
        } elseif {[lindex $seg 2] == "q"} {
            lappend a "q"
        } elseif {[lindex $seg 2] == "9"} {
            lappend a "r"
        } elseif {[lindex $seg 2] == "s"} {
            lappend a "s"
        } elseif {[lindex $seg 2] == "t" | [lindex $seg 2] == "th"} {
            lappend a "t"
        } elseif {[lindex $seg 2] == "U"} {
            lappend a "u"
        } elseif {[lindex $seg 2] == "w"} {
            lappend a "w"
        } elseif {[lindex $seg 2] == "j"} {
            lappend a "y"
        } elseif {[lindex $seg 2] == "z"} {
            lappend a "z"
        } elseif {[lindex $seg 2] == "N"} {
            lappend a "ng"
        } elseif {[lindex $seg 2] == "n~"} {
            lappend a "ny"
        } elseif {[lindex $seg 2] == "aI"} {
            lappend a "ai"
        } elseif {[lindex $seg 2] == "oU"} {
            lappend a "ou"
        } elseif {[lindex $seg 2] == "aU"} {
            lappend a "au"
        } elseif {[lindex $seg 2] == "ia"} {
            lappend a "ia"
        } elseif {[lindex $seg 2] == "uI"} {
            lappend a "ui"
        } else {
            puts ".garbage"
        }
    }
puts ""
puts ""
puts "Target word: $target_word"
```

```
#convert list element to string for comparison
set ajoin [join $a {}]
puts "Transcription: $ajoin "
puts ""

#compare the target and hypo and display result
if {[string match $target_word $ajoin] == 1} {
      puts ":: Correct"
} elseif {[string match $target_word $ajoin] == 0} {
      set fh [open $filename r]
      set data [read $fh]
      close $fh
      foreach a $data {
            if {[string match $a $ajoin] == 1} {
                  puts ":: Correct"
            }
      }
} else {
      puts ":: Miscue"
}

nuke $nnet $wave $stateNet
nuke $feat $observations $prob $probG $vObj
```

Appendix E

CSLU Toolkit's Training Instructions

Source: Adapted from http://www.cslu.cse.ogi.edu/tutordemos/ (Retrieved July 7, 2007)

Training Hidden Markov Model/Aritificial Neural Network (HMM/ANN) hybrids for automatic speech recognition (ASR)

Overall Procedure

Given the background described in the previous section, the process of training a recognizer becomes relatively simple. This section gives the "recipe" for this training process.

1. Create Descriptions

The first step is to create a description of the recognizer and describe how the data will be selected for training. The files that need to be created are:

corpora file

Create a `corpora.txt`. The file contains a master list of each corpus, and the location and format of the files in that corpus. There is no automated way of generating this file, but it is easy to modify by hand. The same corpora file can be used for all training tasks. Thus, the `corpora.txt` file used in this study is constructed as such:

```
corpus: bmwords
     wav_path    /data/speechfiles
     txt_path    /data/txtfiles
     phn_path    /data/phnfiles
     format      {DC-([0-9]+)\.[A-Za-z0-9_]+}
     wav_ext     wav
```

```
txt_ext     txt
phn_ext     phn
cat_ext     cat
ID:         {regexp $format $filename filematch ID}
```

info files

Create "info" files for training, development, and testing. These info files must be created by hand. An info file contains all of the information that is necessary to find examples for training, development, or testing. This info file includes the partition (train, develop, test), how to select the data for the required partition (i.e. filtering parameters), the basename of the recognizer, the minimum number of examples requested for each category, and corpus-dependent information. One info file is required for each of the tasks of training, re-training using forced alignment, development, and testing. The info files used in the study are as follows ('words' is the basename of the recognizer):

<u>words.train.info</u>

```
basename:        words;
partition:       train;
sampling_freq:   16000;
frame_size:      10;
min_sample:      100;

corpus: name:       bmwords
        cat_path:   bmwords_train
        require:    wpt
        filter:     1+1
```

```
lexicon:   words.lexicon
want:      ALL
partition: "{expr $ID % 5} {0 1 2}";
```

words.dev.info

```
partition:  dev;
basename:   words;

corpus: name:           bmwords
        require:        wt
        filter:         1+1
        lexicon:        words.lexicon
        partition:      "{expr $ID % 5} {3}";
```

words.test.info

```
partition:  test;
basename:   words;

corpus: name:     bmwords
     require:     wt
     filter:      1+1
     lexicon:     words.lexicon
     partition:   "{expr $ID % 5} {4}";
```

grammar file

Create a "grammar" file that specifies the grammar that will be used to recognize words. The format of a grammar file is a modification of the ABNF format published by the W3C. The grammar used in this study follows that of a discrete recognition that gives `$grammar = [*sil%%] $word [*sil%%]` where `[*sil%%]` refers to garbage or silence and `$word` takes the word spoken.

`words.grammar`

```
$word = abah | abang | apa | aku | ayat | baca |
        barang | bawang | belalang | betul | bunga | cantik
|
        cendawan | ceria | hampar | jangan | jangkitan |
        kecundang | kelapa | kemarau | kumpulan | makan |
        maklumat | nyata | orang | pandai | pangsa |
        pendatang | pergi | pernah | sayang |
        selesa | suka | udang | umur | wad ;

$grammar = [*sil%%] $word [*sil%%];
```

lexicon file

Create a "lexicon" file that specifies the pronunciation of each word in the grammar. The lexicon file, which is the lexical model constructed with phoneme refinement and treats mispronunciations as alternative pronunciation for an improved accuracy is as follows:

`words.lexicon`

```
abah = A|U (bc bh)|(dc dh)|(pc ph)|m A h;
abang = A (bc bh)|m|(dc dh)|(pc ph) A N;
```

```
apa = (A (dc d)|(pc ph) A|&)|(bc bh A pc ph A);

aku = A kc kh U;

ayat = (A j A tc t|h)|(A j A);

baca = bc bh A|& tS|dZ A|&;

barang = bc bh A|I 9 A|I N;

bawang = bc bh A|& w A N;

belalang = bc bh & l A l A N;

betul = bc bh & tc th o l;

bunga = bc bh U N|n A|&;

cantik = tS|dZ A n tc th E kc k;

cendawan = tS|dZ &|I n dc d A w A n;

ceria = tS &|U 9 ia;

hampar = h A m pc ph A|I 9|(tc t);

jangan = dZ|tS A N|n A n;

jangkitan = dZ|tS A N|n (kc kh)|(gc g) I|E tc th A|I n;

kecundang = kc kh & tS|dZ U n dc d A N|n;

kelapa = kc kh & l A pc ph A|&;

kemarau = kc kh & m A 9 aU|aI|oU;

kumpulan = (kc kh U|& m pc ph U l A n)|(kc kh U pc ph U l
A n);

makan = m A kc k A n;

maklumat = (m A kc k l U m A tc t)|(m A l U m A tc t);

nyata = n~ A tc th A|I;

orang = o 9 A N;

pandai = pc ph A n (dc d)|(tc th) aI;

pangsa = pc ph A N|n s A;

pendatang = (pc ph)|m & n dc d A tc th A N;

pergi = pc ph & 9 gc g I;

pernah = pc ph & 9 n|(tc th) A h;

sayang = s A j A N;

selesa = s & l E|A s A;
```

```
suka = s U kc kh A|&;

udang = U dc d A N;

umur = U m|(pc ph) o|U 9;

wad = w A (dc d)|n;


*sil = (.pau | .garbage);
```

parts file

Create a "parts" file, which specifies how many parts to split each phoneme into, and what context clusters to use. Once again, this must be created by hand.

words.parts

```
A       3 ;
U       3 ;
I       3 ;
i:      3 ;
pc      1 ;
ph      1 ;
&       3 ;
dc      1 ;
dh      1 ;
d       1 ;
tS      1 ;
bc      1 ;
bh      1 ;
E       3 ;
dZ      1 ;
l       1 ;
s       1 ;
```

```
n          1 ;
j          2 ;
th         1 ;
N          2 ;
g          1 ;
gc         1 ;
aI         3 ;
9          2 ;
w          2 ;
m          1 ;
h          1 ;
k          1 ;
kh         1 ;
o          3 ;
oU         3 ;
aU         3 ;
t          1 ;
tc         1 ;
kc         1 ;
n~         2 ;
ia         3 ;
.pau       1 ;
.garbage   1 ;

$sil   = .pau uc .garbage /BOU /EOU ;
$fnt = I i: E A ;
$mid = & ;
$bck = U o ;
$dip = aI aU ia oU ;
$dig = N n~ ;
```

```
$con = t kh k h m g th n s l dZ bh tS dh d ph ;

$sem-vow = w j ;

$vib = 9 ;

$bst_clo = tc kc gc dc bc pc ;
```

2. Find Data

Given the files created above, the scripts to use in order to find data files for training are:

find_files.tcl

Use `find_files.tcl` to find files for training, development, and testing. This script must be called once for each set of files. At this stage, any filters are applied and the corpus is searched for files that are appropriate for the given partition (such as training or testing). When executed, this command gives a list of files in a particular dataset according to the requirement specified in the info files. Figure C.1 depicts a snippet of the output generated. The files are automatically grouped into their corresponding dataset: `words.train.bmwords.files`, `words.dev.bmwords.files`, and `words.test.bmwords.files`.

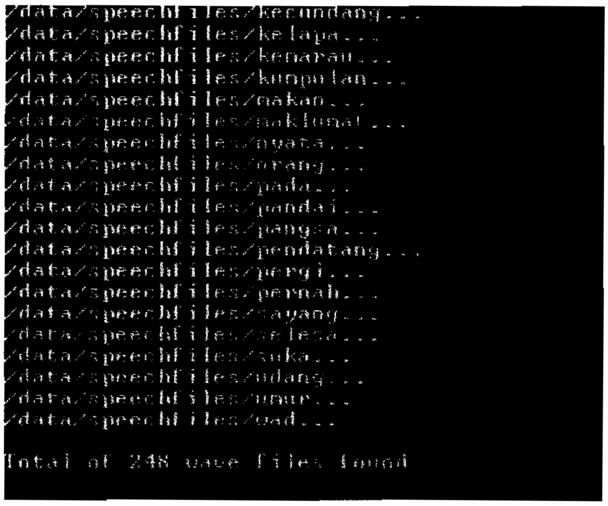


Figure C.1. The output displayed on screen when `find_files.tcl` is executed on `words.train.bmwords.files`.

gen_spec.tcl

Use `gen_spec.tcl` to generate a specification file that contains a list of the categories to train on. This script uses the info, grammar, lexicon, and parts files to create a "spec" file. The specification file contains, in addition to the categories used by the recognizer for training and recognition, the specific frame size, sampling rate, the location of code used to compute acoustic features, the context clusters, and any phonetic mappings.

```
// .spec file
// created Thu Apr 23 07:07:06 Malay Peninsula Standard Time 2009
duration_model    minmax_penalty ;
sampling_freq     16000 ;
frame_size        10 ;


// context clusters
$bck     := U o ;
$bst_clo := tc kc gc dc bc pc ;
$con     := t kh k h m g th n s l dZ bh tS dh d ph ;
$dig     := N n~ ;
$dip     := aI aU ia oU ;
$fnt     := I i: E A ;
$mid     := & ;
```

Figure C.2. The spec file created when running `gen_spec.tcl` extracted from `words.orig.spec`.

gen_catfiles.tcl

Use `gen_catfiles.tcl` to create time-aligned categories from text transcriptions or from phonetic time-aligned transcriptions. These categories are written to separate files with the extension ".cat", which are put in sub-directories that mirror the directory structure of the corpus (or corpora) being used.

revise_spec.tcl

Use `revise_spec.tcl` to (a) tie categories that don't have enough training examples to categories that do have sufficient examples, and (b) update the minimum and maximum duration parameters for each category. `gen_catfiles.tcl` creates output files that indicate the number of examples available for each category, as well as the duration information. The output of this script is a modified "spec" file.

```
// .spec file
// created Thu Apr 23 07:16:59 Malay Peninsula Standard Time 2009
duration_model    minmax_penalty ;
sampling_freq     16000 ;
frame_size        10 ;


// context clusters
$sil     := .pau uc .garbage /BOU /EOU ;
$fnt     := I i: E A ;
$dig     := N n~ ;
$con     := t kh k h m g th n s l dZ bh tS dh d ph ;
$bck     := U o ;
$vib     := 9 ;
$dip     := aI aU ia oU ;
$sem-vow := w j ;
```

Figure C.3. A snippet of the revised spec file extracted from `words.train.spec`.

3. Select Data for Training

Once the files have been selected, the category files have been created, and the description file is correct, then we can use the following scripts and programs to select frames for training:

pick_examples.tcl

Use `pick_examples.tcl` to select examples to train on. The output of this script is an "examples" file, which is used directly by the next script, `gen_examples.tcl`. The examples file contains the selected frames on each of the speech files to be trained on, noted in time in millisecond (ms). A snippet of the example file created is depicted in Figure C.4.

```
/data/speechfiles/bunga/DC-1.bunga5.wav

24 81

24 76

24 82

...

/data/speechfiles/pergi/DC-5.pergi4.wav

10 37

22 50

2 17

...
```

Figure C.4. The examples from speech files selected for training.

gen_examples.tcl

Use `gen_examples.tcl` to create acoustic feature vectors and their associated category information, for each frame to be trained on. This script creates a binary file with the extension ".vec" (for vectors of features).

checkvec.exe

Use checkvec to make sure that the data in the ".vec" file are valid. This program also prints out the number of categories and the number of examples of each category. Figure C.5 represents a snippet of the vector file for train dataset. The number of categories is needed when running nntrain.exe.

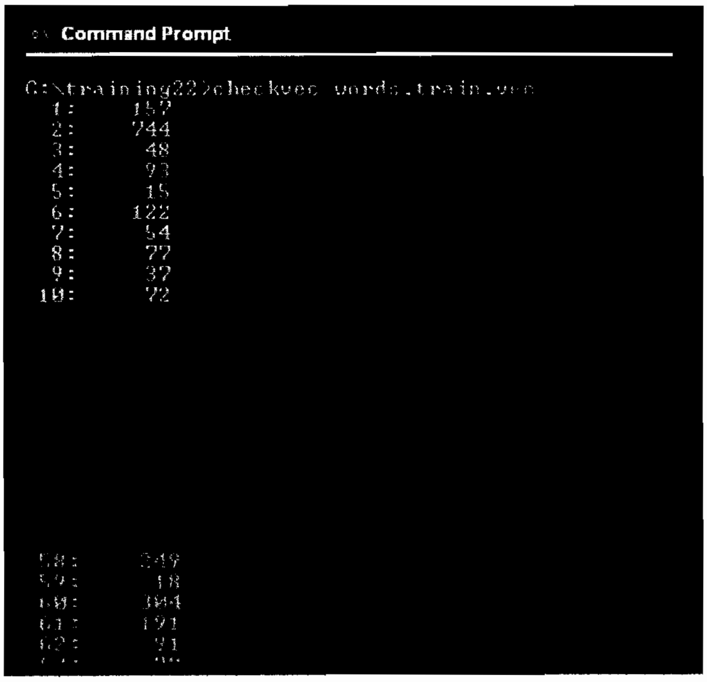


Figure C.5. The vector file created extracted from words.train.vec.

4. Train and Evaluate

nntrain.exe

Use nntrain.exe to train the neural network iterations using the vector file as training data.

select_best.tcl

Use select_best.tcl to find the best iteration of the network using the set of development files.

Note: For output of nntrain.exe and select_best.tcl, please refer to Figure 4.11 of Chapter 4 and Section 5.1 of Chapter 5 respectively.

5. Re-Train

Create force-aligned data using the best iteration of the network that was just trained. To do this, create an info file for forced alignment that specifies a new directory in which to put the category files and a forced-alignment script to use to create the new .cat files. The new info file for force alignment is named words.trainfa.info (see Figure C.6).

words.trainfa.info

```
basename:          words;
partition:         trainfa;
sampling_freq:     16000;
frame_size:        10;
min_sample:        100;

corpus: name:        bmwords
        cat_path:    bmwords_trainfa
        require:     wt
        filter:      1+1
        lexicon:     words.lexicon
        want:        ALL
```

```
force_cat: "fa.tcl wordsnet.26 words.train.spec
           words.lexicon
           WAV TXT c OUT"
partition: "{expr $ID % 5} {0 1 2}";
```

Note that the partition and category path are different than that defined in `words.train.info`. Also, the info file for force alignment contain an extra field, i.e. "force_cat" that specifies using fa.tcl script to generate time-aligned phoneme labels automatically. Hence, the phoneme files are not required. So in the "require" field, only the speech files (w) and the text files (t) are specified.

Then use `find_files.tcl`, `gen_spec.tcl`, `gen_catfiles.tcl`, and `revise_spec.tcl` to generate the force-aligned labels and create a new ".spec" file. Then repeat Step 3 and Step 4 to create a network trained on this force-aligned data.

6. Evaluate Test Set

`select_best.tcl`

Use `select_best.tcl` to evaluate the final best network's performance on the test set. These are the final results that are acceptable for publication.

`asr.tcl`

Use `asr.tcl` to recognize a speech (in this case, read speech recorded) and outputs whether the recognized word (i.e. the speech) is correct or a miscue. This is to evaluate the recognizer's miscue detection rate.

Note: The evaluation is discussed in details in Chapter 5.

Appendix F

Target words and hypotheses (incorrect recognition).

Dataset: Development (words.dev.bmwords.files).

Target	*Hypothesis*	*Edit distance*
abah	buna	4
abang	udang	2
ada	kemarou	5
apa	bapa	1
baca	pada	2
barang	apa	4
betul	pertah	4
betul	upur	4
bunga	baca	3
cantik	hampir	4
cendawan	cindawan	1
cendawan	udang	6
jangan	cindawan	5
jangan	pandai	3
jangan	cendawan	5
jangkitan	hapar	7
jangkitan	janggitan	1
kemarau	barang	5
kemarau	bapa	5
maklumat	malumat	1

nyata	bapa	3
orang	barang	2
pendatang	nyata	5
pernah	abah	4
pernah	pangsa	5
suka	apa	3

Note:

Consideration is given to the letters in the targets and hypotheses with the same place of articulation or manner of articulation. Letter substitutions that do not share either of these are not considered to be included in the phoneme refinement rule. Letters in both targets and hypotheses marked in red to show example of substitutions that are not considered because they do not belong to the same group of either place of articulation or manner of articulation.

Appendix G

Place and Manner of Articulation for BM

(source: Nooraini and Kamaruzaman, 2008)

Manner of articulation	**Place of articulation**					
	Labial	**Alveolar**	**Palate-Alveolar**	**Palatal**	**Velar**	**Glottal**
Plossive-Voiceless	p	t			k	
Plossive-Voiced	b	d			g	
Fricative-Voiceless	f	s			x	h
Fricative-Voiced	v	z				
Affricate-Voiceless			c			
Affricate-Voiced			j			
Nasal	m	n		ny	ng, nx	
Roll		r				
Lateral		l				
Semivowel	w			y		

Appendix H

The Alignment Technique

Recognizer: Lexical model **with** pronunciation variations and phoneme refinement

(After second cycle of force alignment, `wordsfa2net.28`)

Alignment of transcripts, targets, and hypotheses where

- transcripts – human transcription of speech files
- targets – the target words that the dyslexic children are supposed to read aloud
- hypotheses – the recognized speech

Transcript	Target	Hypothesis	
adah	abah	pansa	md
abu	abah	aku	md
abom	abang	pantai	md
abang	abang	abang	
adang	abang	amang	md
aku	aku	aku	
aku	aku	aku	
aku	aku	aku	
apa	apa	apa	
apa	apa	apa	
pe	apa	apa	mud
ayah	ayat	aya	md
ayah	ayat	malumat	md
ayah	ayat	aya	md
baca	baca	barang	fa

baca	baca	baca	
baja	baca	pada	md
barang	barang	barang	
barang	barang	barang	
bewang	bawang	belalang	
bawa	bawang	aku	md
bawang	bawang	bawang	
belalang	belalang	betol	fa
belalah	belalang	abah	md
belalang	belalang	belalang	
betol	betul/betol	betol	
betul	betul/betol	betol	
bunga	bunga	bunga	
buna	bunga	bunga	mud
bena	bunga	abah	md
jantek	cantik/cantek	jangan	md
natik	cantik	pergi	md
cantek	cantik/cantek	cantek	
cedawai	cendawan	cendawan	mud
cindawan	cendawan	cindawan	md
ceda	ceria	selasa	md
ceria	ceria	ceria	
ceria	ceria	ceria	
hemper	hampar	pernah	md
ampar	hampar	barang	md
jemput	jangan	hampat	md
janan	jangan	pandai	md

janan	jangan	jangan	mud
jemtikkan	jangkitan	sayang	md
jangkit	jangkitan	pergi	md
pencudang	kecundang	pansa	md
kecundan	kecundang	kecundan	md
kelapa	kelapa	kelapa	
kelapa	kelapa	kelapa	
kelapa	kelapa	kelapa	
kemeran	kemarau	orang	md
kemaru	kemarau	kemarau	mud
kemurangan	kemarau	kempulan	md
kamponglan	kumpulan	belalang	md
kepulau	kumpulan	kempulan	md
kupulan	kumpulan	kempulan	md
makan	makan	makan	
makan	makan	makan	
makan	makan	makan	
kelumat	maklumat	abah	md
keluwap	maklumat	betol	md
malumat	maklumat	malumat	md
nyata	nyata	nyata	
nyata	nyata	nyata	
nyata	nyata	apa	fa
orang	orang	udang	fa
orang	orang	orang	
orang	orang	orang	
pendendang	pandai	amang	md

andai	pandai	pandai	mud
pandai	pandai	pandai	
pansa	pangsa	pansa	md
pangsa	pangsa	pangsa	
dedate	pendatang	betol	md
pendatai	pendatang	pendatang	mud
pandatang	pendatang	pendatang	mud
berdi	pergi	belalang	md
pegi	pergi	pergi	mud
pergi	pergi	pergi	
pernah	pernah	pernah	
penah	pernah	aya	md
pena	pernah	kelapa	md
sayang	sayang	sayang	
sayang	sayang	sayang	
sayang	sayang	sayang	
selesai	selesa	suka	md
selesai	selesa	selasa	md
selesa	selesa	selesa	
suka	suka	suka	
suka	suka	suka	
suka	suka	suka	
udang	udang	amang	fa
udang	udang	udang	
udang	udang	udang	
umor	umur/umor	aku	fa
umor	umur/umor	wad	fa

umur	umur	upor	fa
uwan	wad	buna	md
wap	wad	bapa	md
wam	wad	aya	md

Note: md – miscue detected

mud – miscue undetected

fa – false alarm

Total correct readings: 48

Total incorrect readings (miscues): 52

Total number of words: 100

Miscue detected, md = 42

Miscue undetected, mud = 10

False alarm, fa = 8

Therefore,

Miscue detection rate = no. of miscue detected / total no. of miscues made

$= (42/52) \times 100\%$

$= 80.77\%$

False alarm rate = no. of correct readings recognized as incorrect / total no. of correct readings

$= (8/48) \times 100\%$

$= 16.67\%$

Appendix I

The Alignment Technique

Recognizer: Lexical model **without** pronunciation variations and phoneme refinement

`(wordsfanet.10)`

Alignment of transcripts, targets, and hypotheses where

- transcripts – human transcription of speech files
- targets – the target words that the dyslexic children are supposed to read aloud
- hypotheses – the recognized speech

Transcript	Target	Hypothesis	
adah	abah	pada	md
abu	abah	aku	md
abom	abang	janan	md
abang	abang	abah	fa
adang	abang	hapar	md
aku	aku	jangan	fa
aku	aku	aku	
aku	aku	aku	
apa	apa	hapar	fa
apa	apa	apa	
pe	apa	pantai	md
ayah	ayat	hapar	md
ayah	ayat	aya	md
ayah	ayat	sayang	md
baca	baca	pandai	fa

baca	baca	baca	
baja	baca	pada	md
barang	barang	barang	
barang	barang	barang	
bewang	bawang	bapa	md
bawa	bawang	bawang	mud
bawang	bawang	bawang	
belalang	belalang	betol	fa
belalah	belalang	wad	md
belalang	belalang	belalang	
betol	betul/betol	pernah	fa
betul	betul/betol	pertah	fa
bunga	bunga	bunga	
buna	bunga	barang	md
bena	bunga	abah	md
jantek	cantik/cantek	pergi	md
natik	cantik	wad	md
cantek	cantik/cantek	cantek	
cedawai	cendawan	cendawan	mud
cindawan	cendawan	cindawan	md
ceda	ceria	selasa	md
ceria	ceria	ceria	
ceria	ceria	selesa	fa
hemper	hampar	jangan	md
ampar	hampar	hampat	md
jemput	jangan	jangketan	md
janan	jangan	pandai	md

janan	jangan	pandai	md
jemtikkan	jangkitan	jangan	md
jangkit	jangkitan	pandai	md
pencudang	kecundang	pernah	md
kecundan	kecundang	kecundan	md
kelapa	kelapa	kelapa	
kelapa	kelapa	bapa	fa
kelapa	kelapa	kelapa	
kemeran	kemarau	suke	md
kemaru	kemarau	kemarau	mud
kemurangan	kemarau	kempulan	md
kamponglan	kumpulan	umor	md
kepulau	kumpulan	kempulan	md
kupulan	kumpulan	hapar	md
makan	makan	makan	
makan	makan	jangan	fa
makan	makan	makan	
kelumat	maklumat	malumat	md
keluwap	maklumat	betol	md
malumat	maklumat	malumat	md
nyata	nyata	nyata	
nyata	nyata	pada	fa
nyata	nyata	makan	fa
orang	orang	orang	
orang	orang	orang	
orang	orang	orang	
pendendang	pandai	hapar	

andai	pandai	pandai	mud
pandai	pandai	pandai	
pansa	pangsa	pansa	md
pangsa	pangsa	pangsa	
dedate	pendatang	betol	md
pendatai	pendatang	pendatang	mud
pandatang	pendatang	pendatang	mud
berdi	pergi	hampir	md
pegi	pergi	pergi	mud
pergi	pergi	pergi	
pernah	pernah	baca	fa
penah	pernah	ayat	md
pena	pernah	pernah	mud
sayang	sayang	baring	fa
sayang	sayang	sayang	
sayang	sayang	sayang	
selesai	selesa	orang	md
selesai	selesa	kempulan	md
selesa	selesa	selesa	
suka	suka	suka	
suka	suka	suka	
suka	suka	suka	
udang	udang	amang	fa
udang	udang	udang	
udang	udang	udang	
umor	umur/umor	umor	
umor	umur/umor	aya	fa

umur	umur	hapar	fa
uwan	wad	bewang	md
wap	wad	abah	md
wam	wad	buna	md

Note: md – miscue detected

mud – miscue undetected

fa – false alarm

Total correct readings: 48

Total incorrect readings (miscues): 52

Total number of words: 100

Miscue detected, md = 43

Miscue undetected, mud = 9

False alarm, fa = 17

Therefore,

Miscue detection rate = no. of miscue detected / total no. of miscues made

= (43/52) × 100%

= 82.69%

False alarm rate = no. of correct readings recognized as incorrect / total no. of correct readings

= (17/48) × 100%

= 35.42%

Appendix J

The Alignment Technique

1. Recognizer: Lexical model **with** pronunciation variations and phoneme refinement

(After first cycle of force alignment, `wordsfanet.22`)

Alignment of transcripts, targets, and hypotheses where

- transcripts – human transcription of speech files
- targets – the target words that the dyslexic children are supposed to read aloud
- hypotheses – the recognized speech

Transcript	Target	Hypothesis	
adah	abah	ada	md
abu	abah	aku	md
abom	abang	jangetan	md
abang	abang	abang	
adang	abang	wan	md
aku	aku	aku	
aku	aku	aku	
aku	aku	aku	
apa	apa	apa	
apa	apa	apa	
pe	apa	baje	md
ayah	ayat	apa	md
ayah	ayat	wan	md
ayah	ayat	adah	md
baca	baca	barang	fa

baca	baca	baca	
baja	baca	baca	mud
barang	barang	barang	
barang	barang	wad	fa
bewang	bawang	bawang	mud
bawa	bawang	bawang	mud
bawang	bawang	bawang	
belalang	belalang	betol	fa
belalah	belalang	aya	md
belalang	belalang	belalang	
betol	betul	beja	fa
betol	betul	betol	
bunga	bunga	bunga	
buna	bunga	udang	md
bena	bunga	beja	md
jantek	cantik	biring	md
natik	cantik	jantek	md
cantek	cantik	cantek	
cedawai	cendawan	cendawan	mud
cindawan	cendawan	cindawan	md
ceda	ceria	curia	md
ceria	ceria	ceria	
ceria	ceria	ceria	
hemper	hampar	cangan	md
ampar	hampar	janan	md
jemput	jangan	baca	md
janan	jangan	janan	md

janan	jangan	janan	md
jemtikkan	jangkitan	jangan	md
jangkit	jangkitan	pandai	md
pencudang	kecundang	pernah	md
kecundan	kecundang	kecundan	md
kelapa	kelapa	kelapa	
kelapa	kelapa	bapa	fa
kelapa	kelapa	kelapa	
kemeran	kemarau	ada	md
kemaru	kemarau	kemarau	mud
kemurangan	kemarau	jendawan	md
kamponglan	kumpulan	umur	md
kepulau	kumpulan	kempulan	md
kupulan	kumpulan	makan	md
makan	makan	makan	
makan	makan	amah	fa
makan	makan	makan	
kelumat	maklumat	beje	md
keluwap	maklumat	betol	md
malumat	maklumat	malumat	md
nyata	nyata	nyata	
nyata	nyata	nyata	
nyata	nyata	apa	fa
orang	orang	udang	fa
orang	orang	orang	
orang	orang	orang	
pendendang	pandai	amah	md

andai	pandai	pandai	mud
pandai	pandai	pandai	
pansa	pangsa	pansa	md
pangsa	pangsa	pangsa	
dedate	pendatang	beje	md
pendatai	pendatang	janan	md
pandatang	pendatang	pendatang	mud
berdi	pergi	beje	md
pegi	pergi	pergi	mud
pergi	pergi	pergi	
pernah	pernah	pernah	
penah	pernah	aya	md
pena	pernah	beja	md
sayang	sayang	aya	fa
sayang	sayang	sayang	
sayang	sayang	aya	fa
selesai	selesa	wan	md
selesai	selesa	selasa	md
selesa	selesa	selesa	
suka	suka	suka	
suka	suka	suka	
suka	suka	suka	
udang	udang	amah	fa
udang	udang	adah	fa
udang	udang	udang	
umor	umur	buna	fa
umor	umur	bunge	fa

umur	umur	bawang	fa
uwan	wad	kupulan	md
wap	wad	bapa	md
wam	wad	wad	mud

Note: md – miscue detected

mud – miscue undetected

fa – false alarm

Total correct readings: 48

Total incorrect readings (miscues): 52

Total number of words: 100

Miscue detected, md = 42

Miscue undetected, mud = 10

False alarm, fa = 15

Miscue detection rate = no. of miscue detected / total no. of miscues made

$= (42/52) \times 100\%$

$= 80.77\%$

False alarm rate = no. of correct readings recognized as incorrect / total no. of correct readings

$= (15/48) \times 100\%$

$= 31.25\%$

***Note:** Even the MDR is equal to that of the proposed model this network is not chosen as to represent the recognizer due to its high FAR.*

2. Recognizer: Lexical model **with** pronunciation variations and phoneme refinement

(After third cycle of force alignment, `wordsfa3net.30`)

Transcript	Target	Hypothesis	
adah	abah	ada	md
abu	abah	aku	md
abom	abang	malumat	md
abang	abang	hampat	fa
adang	abang	amang	md
aku	aku	aku	
aku	aku	aku	
aku	aku	aku	
apa	apa	apa	
apa	apa	apa	
pe	apa	apa	mud
ayah	ayat	ayah	md
ayah	ayat	belalang	md
ayah	ayat	aya	md
baca	baca	baca	
baca	baca	baca	
baja	baca	pada	md
barang	barang	barang	
barang	barang	jangan	fa
bewang	bawang	belalang	md
bawa	bawang	betol	md
bawang	bawang	bawang	

belalang	belalang	betol	fa
belalah	belalang	amang	md
belalang	belalang	belalang	
betol	betul	betol	
betul	betul	betol	
bunga	bunga	bunga	
buna	bunga	udang	md
bena	bunga	barang	md
jantek	cantik	baring	md
natik	cantik	pergi	md
cantek	cantik	cantek	
cedawai	cendawan	cendawan	mud
cindawan	cendawan	cindawan	md
ceda	ceria	pendatang	md
ceria	ceria	ceria	
ceria	ceria	ceria	
hemper	hampar	jangan	md
ampar	hampar	pansa	md
jemput	jangan	udang	md
janan	jangan	jankitan	md
janan	jangan	jangan	mud
jemtikkan	jangkitan	janggitan	md
jangkit	jangkitan	pendatang	md
pencudang	kecundang	pansa	md
kecundan	kecundang	kecundan	md
kelapa	kelapa	kelapa	
kelapa	kelapa	bapa	fa

kelapa	kelapa	kelapa	
kemeran	kemarau	orang	md
kemaru	kemarau	kemarou	md
kemurangan	kemarau	cendawan	md
kamponglan	kumpulan	udang	md
kepulau	kumpulan	kempulan	md
kupulan	kumpulan	kempulan	md
makan	makan	makan	
makan	makan	makan	
makan	makan	makan	
kelumat	maklumat	abah	md
keluwap	maklumat	betol	md
malumat	maklumat	malumat	md
nyata	nyata	nyata	
nyata	nyata	nyata	
nyata	nyata	apa	fa
orang	orang	udang	fa
orang	orang	orang	
orang	orang	orang	
pendendang	pandai	amang	md
andai	pandai	ayah	md
pandai	pandai	pandai	
pansa	pangsa	pansa	md
pangsa	pangsa	pangsa	
dedate	pendatang	betol	md
pendatai	pendatang	pernah	md
pandatang	pendatang	pendatang	mud

berdi	pergi	bewang	md
pegi	pergi	pergi	mud
pergi	pergi	pergi	
pernah	pernah	pernah	
penah	pernah	aya	md
pena	pernah	kelapa	md
sayang	sayang	sayang	
sayang	sayang	sayang	
sayang	sayang	sayang	
selesai	selesa	suka	md
selesai	selesa	selasa	md
selesa	selesa	selesa	
suka	suka	suka	
suka	suka	suka	
suka	suka	suka	
udang	udang	wad	fa
udang	udang	udang	
udang	udang	udang	
umor	umur	buna	fa
umor	umur	udang	fa
umur	umur	orang	fa
uwan	wad	udang	md
wap	wad	bapa	md
wam	wad	bawang	md

Note: md – miscue detected

mud – miscue undetected

fa – false alarm

Total correct readings: 48

Total incorrect readings (miscues): 52

Total number of words: 100

Miscue detected, md = 47

Miscue undetected, mud = 5

False alarm, fa = 10

Miscue detection rate = no. of miscue detected / total no. of miscues made

= $(47/52) \times 100\%$

= 90.38%

False alarm rate = no. of correct readings recognized as incorrect / total no. of correct readings

= $(10/48) \times 100\%$

= 20.83%

Note: Even though the MDR is high - 90.38%, the FAR is also relatively higher with 20.83% due to the tradeoff between MDR and FAR. This network is considered not at optimum performance since its WER is relatively higher with 35.85% and does not meet the satisfaction range of 25% to 30% (see Chapter 2).

List of Related Publication to Thesis

Husniza, H., & Zulikha, J. (2010) Improving ASR performance using context-dependent phoneme models. *Journal of Systems and Information Technology (JSIT), 12*(1).

Husniza, H., & Zulikha, J. (2009). Pronunciation variations and context-dependent model to improve ASR performance for dyslexic children's read speech. *Proceedings of International Conference on Computing and Informatics (ICOCI 2009)*, Kuala Lumpur, Malaysia.

Husniza, H., & Zulikha, J. (2009). ASR technology for children with dyslexia: Enabling intermediate intervention to support reading in Bahasa Melayu. *US-China Education Review, 6*(6), 64-70.

Husniza, H. (2009). Automatic speech recognition model for dyslexic children with reading and spelling discrepancy. Poster session presented at the PhD in IT Seminar Abstract and Poster Book. UUM CAS, UUM.

Husniza, H., & Zulikha, J. (2009). Dyslexic children's reading pattern as input for ASR: Data, analysis, and pronunciation model. *Journal of Information and Communication Technology (JICT), 8,* 1-13.

Husniza, H., & Zulikha, J. (2008). A retrospective and future look at speech recognition applications in assisting children with reading difficulties. *Proceedings of the World Congress on Engineering and Computer Science WCECS 2008*, San Francisco, USA, 555-558.

Husniza, H., & Zulikha, J. (2008). Towards an ASR-based application to support reading for dyslexic children: Constructing the pronunciation model. *Proceedings of International Conference on IT and Multimedia, ICIMu 2008*, Bangi, Malaysia.

Husniza, H., & Zulikha, J. (2008). ASR technology for immediate intervention to support reading for dyslexic children. *Proceedings of the International Conference on Education, Research, and Innovation*, Madrid, Spain.

Husniza, H., Zulikha, J., & Fakhrul, A. A. (2008). Support models using automatic speech recognition in automated reading tutor for dyslexic children. *Proceedings of the 3rd International Educational Conference UPSI-UPI 2008*, UPSI, Malaysia.

Fakhrul, A. A., Husniza, H., & Zulikha, J. (2008). Defeating phonological-core deficit among dyslexic children by activating arousal dimension of affective impression. *Proceedings of the International Conference on Education, Research, and Innovation,* Madrid, Spain.

Husniza, H., & Zulikha, J. (2007). How ASR can assist children with dyslexia in literacy skills. *Proceedings of the 1st Regional Conference on Computational Science and Technologies, RCCST 2007*, Kota Kinabalu, Sabah, 176-179.

Husniza, H., & Zulikha, J. (2007). Automatic speech recognition to assist dyslexic children in reading. *Proceedings of Rural ICT Development RICTD 2007*, UUM.