THE DESIGN AND VERIFICATION OF MALAY TEXT TO SPEECH SYNTHESIS SYSTEM

TAN TIAN SWEE

A thesis submitted in fulfillment of the
requirements for the award of the degree of
Master of Engineering (Electrical)

Faculty of Electrical Engineering
Univeriti Teknologi Malaysia

APRIL 2004

Dedicated to Jesus Christ,
my personal Savior and Lord,
my pastor, Church members,
my beloved mum, dad, sister, and brother.

ACKNOWLEDGEMENTS

There are a very large number of people to thank in connection with this work. I shall begin at the beginning, by thanking my original supervisor, the Associate Professor Dr. Sheikh Hussain Shaikh Salleh. . To him I am deeply grateful, for his valuable guidance, advice and suggestion. To me, he is more like my friend than my supervisor that continuously encourage me. Thanks to him for giving me the flexibility in doing the research.

A special thanks to Associate Professor Dr. Aini Hussain for her many comments on earlier versions of this thesis and guidance in developing Malay TTS especially the NLP modules. Thanks to go Ng Ming Kwang, my Church members, and my pastor, Pastor Ong for their moral support and their help in carrying the questionnaire survey at the end of this project. My sincere thanks especially for the technical and emotional support from DSP-lab: Helmi, Alwi, Hong, Mala, Prakesh, Zul, Zamri, Jeffry etc. Special Thanks to Dr Allan Black from Carnergie Mellon University, who had provided me with information and advises related to my research.

Most of all, thanksgiving, praise and glory is all to Jesus Christ, who gives me grace, love, patience, healthy, wisdom and ability to walk through all the problems and obstacles during the period of my study.

ABSTRACT

Synthetic or artificial speech has been developed steadily during the last decades. The intelligibility of synthetic speech has reached an adequate level for most applications, especially for communication impaired people. The first objective of this work is to design and develop a Malay Text to Speech (Malay TTS) system. This will include the design of Malay TTS diphone database, tokenization rules, letter-to-sound rules, Malay lexicon and prosody rules. Other focus of this work is to design a set of test methods specifically for verifying Malay TTS performance. This work has produced a diphone database with 1629 diphone file in residual-exited LPC (RELP) format and its total size is around 3.4 Mega bytes. Besides that, this work also has identify the possible tokenization area in Malay TTS and develop a digit tokenization for Malay TTS as the basic for further development of more complete tokenization rules. This work also has produced complete letter-to-sound (LTS) rules for Malay primary word that has high accuracy and almost 100 percent accuracy. A set of lexicon containing 1000 most common use Malay words also being setup as complement to the LTS coverage. A set of a prosody rules using a CART tree has been setup as the preliminary study in prosody design for Malay TTS. Finally, the very first try in designing the testing methods and procedures for Malay TTS has been completed. It will provide a more complete technique in verifying the performance of Malay TTS that will become the benchmark for Malay TTS evaluation and improvement in future.

ABSTRAK

Ucapan sintetik telah berkembang pesat sejak beberapa dekad kebelakangan ini. Kebolehdengaran ucapan sintetik sudah mencapai satu tahap yang tinggi dalam penggunaannya dalam pelbagai bidang terutamanya untuk komunikasi bagi orang yang kehilangan keupayaan. Objektif pertama bagi kerja ini ialah merekabentuk dan menghasilkan satu sistem Sintetik Ucapan Melayu. Ini termasuk merekabentuk dan menghasilkan pangkalan data diphone, peraturan untuk menukar tanda (nombor atau simbol) kepada perkataan (tokenization rules), peraturan penukaran huruf kepada suara (letter-to-sound rules), pangkalan data yang mengandungi perincian ejaan perkataan (Malay lexicon), dan peraturan persajakan (prosody rules). Fokus lain ialah merekabentuk dan menghasilkan satu set pengujian untuk mengenal pasti kualiti sistem Sintetik Ucapan Melayu. Projek ini telah menghasilkan satu pangkalan data diphone sebesar 3.4 Mega bait yang mengandungi 1629 diphone. Selain itu, projek ini juga telah mengenal pasti semua kemungkinan tanda dalam Bahasa Melayu dan merekabentuk penukaran digit ke perkataan sebagai asas kepada kajian penukaran tanda. Projek ini telah menghasilkan peraturan penukaran huruf kepada suara yang lengkap untuk perkataan asas Bahasa Melayu dan mempunyai ketepatan yang hampir 100 peratus. Satu pangkalan data ejaan perkataan sebanyak 1000 perkataan sering digunakan juga telah direka untuk melengkapi peraturan penukaran huruf kepada perkataan. Akhir sekali, peraturan persajakan dengan **CART** xxtelah direka sebagai langkah pertama dalam kajian persajakan. Akhir sekali, projek ini telah berjaya merekabentuk satu set ujian untuk Sintetik Ucapan Melayu buat kali pertama dan ini akan menjadi tanda aras yang akan digunakan untuk mengkaji kualiti Sintetik Ucapan Melayu.

TABLE OF CONTENT

CHAPTER	TITLE	PAGE
	TITLE PAGE	i
	ADMISSION PAGE	ii
	DEDICATION	iii
	ACKNOWLEDGEMENT	iv
	ABSTRACT	v
	ABSTRAK	vi
	TABLE OF CONTENTS	vii
	LIST OF TABLES	xi
	LIST OF FIGURES	xiii
	LIST OF SYMBOLS	xv
	LIST OF APPENDIXES	xvii

THESIS CONTENT

CHAPTER 1	**INTRODUCTION**		**1**
	1.0	Background of research	1
	1.1	Overview of Malay Text To Speech System	2
	1.2	Objective	4
	1.3	Scopes of the Project	4
	1.4	Research Methodology	5
	1.5	Thesis layout	5
	1.6	Contribution of the Thesis	7

CHAPTER 2 REVIEW OF TEXT TO SPEECH SYSTEM 9

2.0 Introduction 9

2.1 Overview of Text To Speech System 9

2.1.1 History of Synthesis Techniques 10

2.1.2 History of Malay Text to Speech in Malaysia 12

2.2 The Human Speech Production System and Speech Wave Generation Methods 13

2.2.1 The Human Speech Generation 13

2.2.2 The Source-Filter Theory of Speech Production 16

2.3 Synthesis techniques 17

2.3.1 Formant Synthesis 17

2.3.2 Articulatory Synthesis 18

2.3.3 Concatenative Synthesis 19

2.3.3.1 Concatenative methods 20

2.3.3.2 Level of Concatenation 23

2.4 Current Commercial and Non-commercial TTS System 25

2.5 Programming language and Speech Processing Tools 29

2.5.1 Scheme Programming Language 29

2.5.2 Software for Speech Processing 29

2.6 Conclusion 31

CHAPTER 3 Database Design For MALAY TEXT TO SPEECH 32

3.0 Introduction 32

3.1 Malay Phonetics and Phone sets 33

3.1.1 Classification of Malay Phonemes 33

3.1.2 Malay Phonemes Feature Analysis 38

3.2 Malay Morphology 39

3.2.1 Malay Stem Word 40

3.2.2 Malay Derivative Word 40

3.2.3 Malay Reduplicate word 43
3.3 Sentence Structure 43
3.4 Diphone Database Design 45
3.4.1 Concatenative process of Malay Diphone 45
3.4.2 Festival Diphone Database Architecture 46
3.4.3 Malay Diphone Database Design 47

CHAPTER 4 DESIGN AND IMPLEMENTATION OF MALAY TEXT TO SPEECH SYSTEM 59

4.0 Introduction 59
4.1 Malay TTS system 59
4.2 Malay TTS NLP Module Generation 63
4.2.1 Common NLP 63
4.2.1.1 Tokenization 63
4.2.1.2 Letter-To-Sound (LTS) Module and Lexicon 64
4.2.1.3 Prosody 65
4.2.2 Malay TTS NLP Design 68
4.2.2.1 Tokenization 68
4.2.2.2 Building the Lexicon and LTS 73
4.2.2.3 Add Prosody 88
4.3 Malay Waveform Generator Modules Generation 91
4.4 Malay TTS system 92
4.5 Testing Method Design 92
4.5.1 Testing and Evaluation 93
4.5.2 Pre-Test 95
4.5.3 Intelligibility Test 95

CHAPTER 5 RESULT AND DISCUSSION 101

5.1 Results 101
5.1.1 Testing and Evaluation 104

5.1.2 Perceptual 111
5.2 Conclusion 112

CHAPTER 6 CONCLUSION AND SUGGESTION 114
6 Introduction 114
6.1 Conclusion 114
6.2 Suggestion for Future Development 115
6.2.1 Improvement of Database Processing Method 115
6.2.2 Improvement of LTS method 116
6.2.3 Improvement of Waveform Generation Synthesis Method. 116
6.2.4 Improvement in Prosody. 117
6.2.5 Malay TTS Application 117

REFERENCES 118

APPENDIXES
APPENDIX A-H 124

LIST OF TABLES

TABLE NO.	TITLE	PAGE
2.1	The comparison between current commercial and non commercial TTS system.	28
3.1	The functions of vowels in Malay language	36
3.2	The functions of consonants in Malay Language	36
3.3	Malay vowel phonemes according to position and height of tongue	38
3.4	Classification of consonants according to manner of articulation, place of articulation, and voicing	39
3.5a	The structure of monosyllabic words in Malay language	41
3.5b	The structure of disyllabic words in Malay language	41
3.5c	The structure of tri-syllabic words in Malay language	41
3.5d	The structure with four syllables or more	41
3.6	The examples of Malay Reduplicate word	43
3.7	The examples of sentences in Malay Language	44
3.8	The structure of a sentence in Malay language	45
3.9	The example of pronunciation of Malay Words	45
3.10	Example of word and sentence form through the combination of diphone	46
3.11	The diphone possibility for Malay language	51
3.12	The diphone pairs example According to its occurrence	53
4.1	Categories of Malay digit and abbreviation together with their examples	69
4.2	Examples of Malay Abbreviations and its pronunciation in Addenda Lexicon	73
4.3	Examples of words and their syllables boundary	

	representation	83
4.4	Malay phonemes and their representation in Malay TTS	85
4.5	The sample of Modified Rhyme Test (MRT) listening test Word List	97
4.6	The Sample of Modified Rhyme Test Analysis Table	97
4.7	The Sample of Result Table for Mean Opinion Score (MOS) Test	98
4.8	The Sample of Mean Opinion Score (MOS) Test Analysis Table for Word 1	99
4.9	The Sample of Mean Opinion Score (MOS) Test Result	100
5.1	General Information for the Participants	103
5.2a	The MRT Test Result for Question 1 to 30	105
5.2b	The MRT Test Result for Question 31 to 50	106
5.3	The MRT Test Error Score According to Participant	108
5.4	The MOS test result	109
5.5	The Visual Perceptual Test table for word “alamat”.	111
5.6	The Visual Perceptual Test table for sentence “nama saya Tan”.	112
5.7	The Rating and Its Equivalent Representation in Term of Level of Performance and Percentage of Accuracy	113
6.1	Examples of different types of pronunciation for same word	116

LIST OF FIGURE

FIGURE NO.	TITLE	PAGE
1.1	The Architecture of Malay TTS synthesizer.	3
1.2	The 5 phases involved for designing Malay TTS system.	6
2.1	Simple Text To Speech synthesis procedure.	10
2.2	The Human Speech Production System.	14
2.3	Cylindrical tube of varying cross sectional area to represent the vocal tract	14
2.4	Parallel and Cascade Configuration of the Formants in Formants Synthesis Method	18
2.5	Block diagrams of Residual-excited Linear Predictive (RELP)	22
2.6	Pitch modification of a voiced speech segment	23
2.7	The Emu_labeler outlook	30
2.8	The Speech Analyzer outlook	30
3.1	US English Phone set	34
3.2	Malay Phone set	34
3.3	Vowels in Malay Language according to the position and height of tongue	35
3.4	Diphthong in Malay Language	37
3.5	The concatenative process of diphone	46
3.6	The Diphone database architecture with direct access method	48
3.7	The lab file for the diphone after hand checking using Emu_labeler	56
3.8	The index listing for the diphone	56
4.1	The Architecture of Malay TTS system	60
4.2	The Process of Malay Text to Speech synthesizing	61
4.3	The Tokenization process	64

4.4	The Architecture of Dictionary LTS Modules	65
4.5	The Architecture of Rule Base LTS Modules	65
4.6	Prosody dependencies	66
4.7	The Malay Tokenization Modules	69
4.8	Flow chart for tokenization of normal number or date to word	71
4.9	The Malay LTS Modules	74
4.10	The process of parsing the input sentence to match its phoneme pattern	75
4.11	Processing of input text with lookup process through lexicon and LTS rules	77
4.12	The syllable groups for primary words	81
4.13	The flow chart for LTS rules setting	82
4.14	The Malay language syllables that has being group according to specific group	84
4.15	The “CC” occurrences in Malay syllables	86
4.16	The Malay Prosody Modules	88
4.17	Prosody Rules for Malay TTS system	90
4.18	The effect of break insertion when punctuation or content word detected	90
4.19	The DSP Architecture of Malay TTS	91
4.20	The outlook of Malay TTS system	92
5.1	The subject in testing	102
5.2	State of Origin for 40 Participants	104
5.3	The Percentage of Participants According to Age Group	104
5.4	Total Error Percentage for MRT Test According to Question	106
5.5	Total Error Percentage for MRT Test According to Participants	107
5.6	Mean Opinion Average Score for Word 1	109
5.7	Mean Opinion Average Score for Word 2	110
5.8	Mean Opinion Average Score For Sentence 1	110
5.9	Mean Opinion Average Score for Sentence 2	110

LIST OF SYMBOLS/ ABBREVIATIONS

An	the cross-sectional area
c	the pressure waves
CHART	Classification and Regression Trees
CSTR	Center for Speech Technology Research
CVV	Consonant-Vowel-Consonant
DM	database module
DSP	Digital Signal Processing
FSS	Festival Speech Synthesis
F0	The Fundamental Frequency
$G(z)$	transfer function of the glottal waveform
	"filter"
HAMLET	Helpful Automatic Machine for Language and
	Emotional Talk
HMM	Hidden Markov Models
$H(z)$	the transfer function of a digital filter
ICT	Information and Communication Technology
IPA	International Phonetic Association
IPS	International Phonetic Symbols
L&H	Lernout & Hauspies
LP	Linear Prediction
LTS	Letter to Sound
LTSM	letter-to-sound rule module
Malay TTS	Malay Text To Speech
MOS	Mean Opinion Score
MRT	Modified Rhyme Test
MSC	Multimedia Super Corridor

NLP	Natural Language Processing
PAT	Parametric Artificial Talker
PSOLA	pitch-synchronous overlap and add
P_0	air density
RELP	Residual Excited LPC
SIOD	Scheme in one Defun
SPM	simple prosody module
S.U.M	Malay Speech Synthesis or in Malay language "Sintesis Ucapan Melayu"
$S(z)$	the z-transform of the speech signal
TTS	text-to-speech
TM	tokenization module
UKM	University Kebangsaan Malaysia
$U(z)$	an approximation to the excitation signal
$V(z)$	transfer function of the vocal tract
VCV	vowel-consonant-vowel
WG	Waveform Generation
$R(z)$	the radiation characteristic
?	IPA symbol for Malay phoneme gh
?	IPA symbol for Malay phoneme kh
?	IPA symbol for Malay phoneme ng
š	IPA symbol for Malay phoneme sy
e	IPA symbol for Malay phoneme e (pepet)
e	IPA symbol for Malay phoneme e (taling)

LIST OF APPENDIX

APPENDIX	TITLE	PAGE
A1	Scheme Programming Language	124
A2	Emu_labeller	133
A3	Speech Processing Tools	135
B1	Malay Baku Pronunciation	137
B2	Recording specification	142
C	LPC ANALYSIS AND SYNTHESIS DETAIL	143
D	Data and results	156
D1	Malay Diphone list with real word carrier	156
D2	Malay Diphone Index File	169
D3	Malay Compile lexicon	179
D4	Malay LTS testing	186
E	Malay TTS application	188
F	Questionnaire and Testing	206
F1	Testing Table	206
F2	Questionnaire	210
F3	Questionnaire slide	214
F4	Questionnaire answer and Evaluation	216

CHAPTER 1

INTRODUCTION

1.0 Background of Research

Since the launching of Multimedia Super Corridor (MSC) project in Malaysia, the Information and Communication Technology (ICT) has been growing rapidly. As a result, computer system as a tool for information and communication medium is becoming more important since then. In addition, the human computer interaction system which involved speech recognition, synthesis etc. also experiences tremendous growth, resulting in many applications being developed and commercialized. For instance, Microsoft recently launched the Office XP that has the capability to pronounce (or read aloud) the text input using the Speech synthesis engine. Indeed, speech synthesis has been very useful in helping human in various areas such as telephone speech, application in cars, public information systems, education assistance tools, email reading etc (Mangold, 2001). The Text to Speech (TTS) system is also useful for the physically handicap. For example, speech synthesis has been used as reading and communication tools for visually impaired. The first commercial TTS system is Kurzweil Reading Machine for the blind introduced by Raymond Kurzweil in the late 1970's (Klatt, 1987). For the hearing impaired and vocally handicapped, the TTS system has been used as a communication tool with people who are sign language illiterate (Gold and Morgan, 2000). Another application of the TTS system is Helpful Automatic Machine for Language and Emotional Talk (HAMLET), which is developed to help users to express their feelings (Lemmetty, 2001).

Unfortunately, all current available commercialize TTS systems are designed in other languages such as English, German, Japanese, Thailand and Chinese etc (Taylor et al, 1999). As for the Malay TTS system is still under research and is yet to be as famous as the others. Therefore, the aim of this project is to build a Malay TTS system as one of the preliminary step towards the development of the first commercialized Malay TTS system. Besides that, it is hope that the outcome of this work, which is the Malay TTS system, will be adopted and applied in the various aforementioned applications. Such use can benefit our country and other Malay speaking countries.

1.1 Overview of Malay Text To Speech System

Text to Speech (TTS) system is a Speech synthesis tools that is able to pronounce any input raw texts aloud (Tan et al, 2003). Basically TTS system is divided into two main components (Tan and Sheikh, 2003).

The block diagram of Malay TTS is shown in Figure 1.1. The first component of TTS system as illustrated in Figure 1.1 is the Natural Language Processing (NLP) module (Donovan, 1996). This component acts as a black box that processes the input raw texts using the linguistic rules that will then assign or output the phoneme and prosody for the input texts. For Malay TTS system, the NLP component consists of three modules: the letter-to-sound (LTS) rules, Tokenization rules, and Prosody Phrasing method such as intonation, phrase break assignment and duration setting (Black et al, 1998). The LTS module consists of a set of rules that associates each letters to its sound according to the Malay language pronunciation rules (Pagel et al, 1998). Meanwhile the Tokenization rules comprises of a set of rules to change token word (e.g. numbers, date, symbols etc.) to full text format such as the digit “100” is replaced with “seratus” and date “10hb Mei 2002” is replaced with “sepuluh haribulan Mei tahun dua ribu dua” (Tan and Sheikh, 2003). Finally, the Prosody Phrasing module involves applying the Malay language intonation rules and the duration pattern to enable the synthesized speech to have the Malay accent (Alan and Kevin, 2000).

The second component of the TTS system (as shown in Figure 1.1) is the waveform generator black box which uses the phoneme and prosody produced by NLP to match the phoneme to the pre-stored sound database and concatenate the phonemes to produce a continue set of waves file or speech sound that contain the prosody features (Taylor et al, 1999). Since this project not focus on building the wave generator, the residual-excited LPC wave generation method (the default wave generation module in Festival) has been chosen as wave generation method for Malay TTS (Macon et al, 1997).

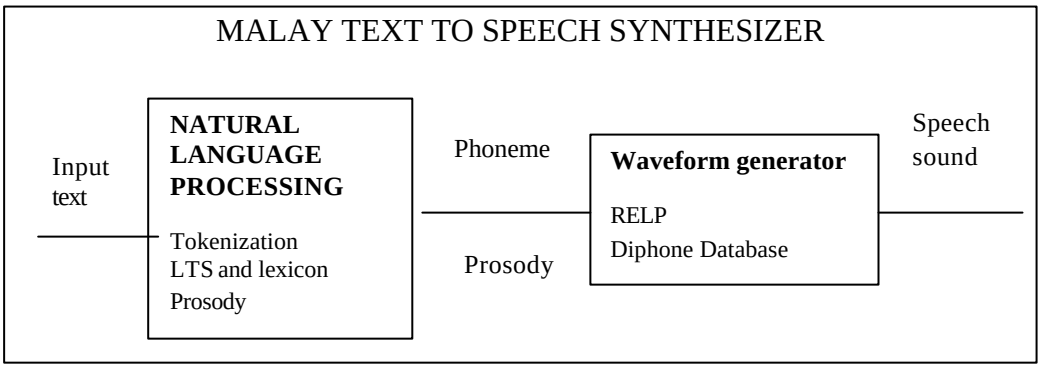

Figure 1.1: The Architecture of Malay TTS synthesizer (Tan and Sheikh, 2003).

Typically in TTS system development, the first and of utmost importance would be the design of the database, which contains the small unit of speech such as phoneme, diphone, syllable, word etc (Gold and Morgan, 2000). But the diphone unit has been chosen for this project because of few specific reasons such as the size of database and it contain more articulate information than phoneme (Donovan, 1996). This will be discussed in details later in Chapter 3. A diphone unit is a combination of two phones such as “a-b”, “m-n”, “t-a” and “s-u” etc (Tan et al, 2003). TTS system using diphone concatenation method is also named as Diphone Synthesis TTS system (Tan and Sheikh, 2003), and this project will produce a Malay Diphone Synthesis TTS system.

1.2 Objective

There are four main objectives to be achieved in this project. The first objective of this project is to build a Malay TTS Engine through Festival Speech Synthesis system that is able to pronounce any input raw text with high accuracy.

Secondly, a survey or detail study of Malay Linguistic to find or design a best database for the Malay Text To Speech Synthesis system in defining the Phone Set, Phone Duration, Diphone List and recording of the database is hope to be realized.

The third objective of this project would be the design of Malay Lexicon, Malay Letter to Sound Rules, Malay Tokenization Rules and simple Malay Prosody.

Finally, a test or evaluation method will be designed carefully that can be used specifically for Malay TTS engine. This would be useful as though still lack of evaluation method for Malay TTS system.

But due to the time limitation, certain modules of the Malay Speech Synthesizer such as Tokenization rules, and prosody will only focus on preliminary study that will be the stepping-stone for further research purpose.

1.3 Scopes of the Project

There were several concerns before defining the scopes of this project. Knowing the fact that the technique in developing TTS system has evolved and matured considerably for the last few decades, a dilemma arose whether to initiate the project from the scratch or to utilize currently available TTS development tool. If the project were to be started from scratch, it would take a very long time before the TTS system can be finalized, commercialized and used for detail and in-depth study. Alternatively, utilizing the available development tool can give a jump-start towards the creation of the TTS system, hence shorten the development time. Therefore, the

Festival speech synthesis system has been selected and used in the development of the Malay TTS system for this project.

By using Festival Speech Synthesis system, the project will concentrate mainly on the design of specific modules relevant to the Malay TTS system such as the database module (DM), letter-to-sound rule module (LTSM), tokenization module (TM) and simple prosody module (SPM). Currently, the NLP component which consists of the LTSM, TM and SPM are available in English and Spanish version. As such, it is vital to develop the NLP component specifically for the Malay language.

1.4 Research Methodology

There are five phases for designing the Malay TTS system as shown in Figure 1.2. The first phase involved the database design and the second phase involved NLP Configuration or Modification. When the NLP and database are ready, Malay TTS will be setup and then it will pass through the testing process to verifying the quality and performance of the system. Finally, it has deployed in two simulations such as Intelligent Security Door and Talking Clock to validate the system (Appendix E). The testing method will be the benchmark as though it is a very first try on designing the testing method and procedure specifically for Malay TTS.

1.5 Thesis layout

This thesis has been divided into five major parts. The first part of which is Chapter 1 has included the introduction, background, objective and scope of the project. It will briefly discuss the main idea and the aim of the project. It will also cover the scopes of the project that reflect the feasibility of this study.

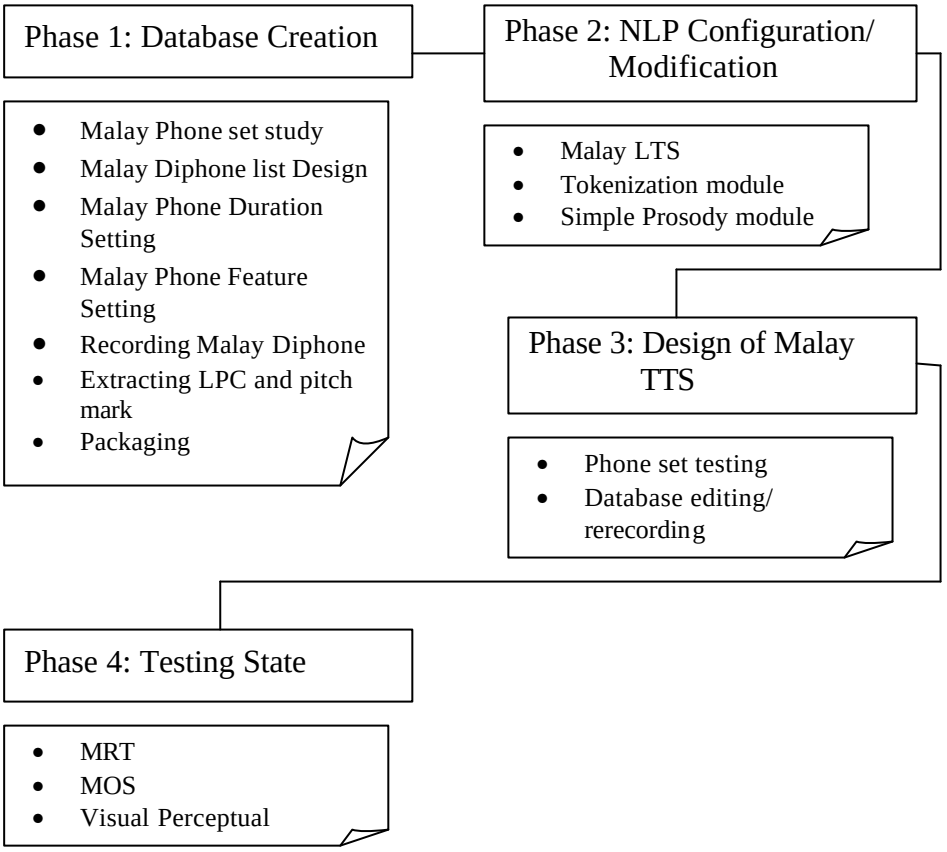

Figure 1.2: The 5 phases involved for designing Malay TTS system.

Chapter 2 will discuss more about of the project background, history and related works in more details. This part will also highlight the important things or methods that would be useful for the preliminary study before starting of the whole project. This part will focus more on the TTS background, history, techniques, and speech processing tools that will be used for developing Malay TTS.

The third part of this thesis is database design in Chapter 3. This chapter will discuss the background of Malay linguistic. These linguistic rules will be applied to diphone database design at the end of Chapter 3 and NLP design in Chapter 4.

The fourth part of this thesis is the about the design and implementation of Malay TTS system which will be elaborated in Chapter 4. This part will discuss in detail on how and what kind of method that would be used for the project. The types of methods employed to evaluate the system and the expected result would be mentioned as well.

The fifth part of this thesis will focus on how to test the whole system that has been developed. Results, which were obtained from the project and the evaluation for the result by comparing to earlier expected result, are discussed. Chapter 5 will include all the experiment, results and evaluation procedure.

The final part of the thesis would be the conclusion and recommendation for the project. Conclusion based on the results obtained through experiments in Chapter 5 would be made. Some recommendations for improvements of the project in the future are being suggested as well.

1.6 Contribution of the Thesis

This project has developed a set of diphone database with 1629 diphone units for Malay TTS system. The research is capable to minimize the database required for Malay Diphone Text To Speech system. For instance, if all the vowel combination (total 10) being putted in the database, the total phoneme plus vowel combination and diphthong will be 42 (32 phoneme and 10 vowel combination). This will result the total combination of diphone units as 42*42 or 1764. This thesis also provide a way for further study in improving Malay TTS's phone set, phone definition and its duration in future.

A set of Malay TTS NLP engines such as Malay LTS modules, Malay tokenization rules, and basic prosody module have been developed. The study does not require the implementation of the whole library in the lexicon module for the Malay TTS system. It only requires 1000 words in lexicon database compare to over 20000 words in lexicon database for English TTS system (Alan and Kevin, 2000). The preliminary study on prosody and the development of basic number to word conversion in tokenization module has provided some basic ideas for future development of more complete Malay TTS system. The LTS module for Malay TTS system has been tested with small conventional Malay dictionary that contains over 4000 most common used Malay words and has almost 100% accuracy after the correction of the mismatch rules.

Finally, this project has produced a set of testing and analysis method and procedure specific for Malay TTS system that will be the benchmark of the testing procedure for future Malay TTS system. The intelligibility accuracy of over 85% has proved its quality in intelligibility test. Malay TTS project also has successfully validated through the implementation of two simulations such as Intelligent Security Door System and Talking Clock (Appendix E).

some lacking also in terms of its naturalness and some artifact. To solve these problems some newer waveform generation method such as PDSOLA and harmonic sinusoidal method can be used to improve its performance (Lemmetty, 1999).

6.2.4 Improvement in Prosody

As this system using simple prosody rules to predict its intonation and duration, it can be improved by improving the simple prosody rules with more rules to predict the intonation of the sentences. Some other prosody method such as TILT Model and ToBI can be used as the research method for the performance of the prosody rules (Black and Hunt, 1996).

6.2.5 Malay TTS Application

The Malay TTS system also can be applied in other areas such as Audio Visual- Talking Head, real security door system, Assistance tools for the blind, and SMS reader (Cox et al, 2000). This would be able to make it more popular in terms of its application and area of usage.

REFERENCE

Alan, B. and Kevin, A. L. (2000). *Building Voice in the Festival Speech Synthesis System: Processes and issues in building speech synthesis voice.* Ed 1.2 beta, US: Carnegie Mellon University.

Andersen, O., Dyhr, N.J., Engberg, I. S., Nielsen, C. (1998). Synthesizing Short Vowels from their Long Counterparts in a Concatenative Based Text-to-Speech System. *3rd ESCA Workshop on Speech Synthesis.* November 26-29. Australia: ESCA, 147-151.

Asmah, H.O. (1983). *The Malay Peoples of Malaysia and Their Languages.* Malaysia: Dewan Bahasa Dan Pustaka.

Black, A. and Campbell, N. (1995). Optimising selection of units from speech databases for concatenative synthesis. *Proceeding of Eurospeech 95.* September. 18-21. Madrid, Spain: Eurospeech, vol 1, 581-584.

Black, A. and Lenzo, K. (2001). Flite: a small fast run-time synthesis engine. *4th Speech Synthesis Workshop.* August 29 - September 1. Scotland: ISCA 204-208

Black, A. and Hunt, A. (1996). Generating FO contours from ToBI labels using linear regression. *Proceedings of ICSLP 96.* October 3-6. Philadelphia, Penn: ICSLP: vol 3, pp 1385-1388.

Black, A., Lenzo, K. and Pagel, V. (1998). Issues in Building General Letter to Sound Rules. *3rd ESCA Workshop on Speech Synthesis.* November 26-29. Jenolan Caves, Australia: ESCA98, 77-80.

Black, A. and Taylor, P. (1997). Automatically clustering similar units for unit selection in speech synthesis. *Proceedings of Eurospeech 97.* September 22-25. Rhodes, Greece: Eurospeech97, 601-604.

Boogart, T. and Silverman, K. (1992). Evaluating the overall comprehensibility of speech synthesizers. *Proceedings of the International Conference on Spoken Language Processing.* August 2-4. Alberta, Canada: ICSLP,1207-1210.

Briony, J. W. (1994). Welsh Letter-to-Sound Rules: Rewrite Rules and Two-Level Rules Compared. *Processing of Computer Speech and Language. vol 8.* CSL94, 261-277.

Bryan, L. P. (1998). *Enhancement, Segmentation, and Synthesis of Speech with Application to robust Speaker Recognition.* Duke University: Ph.D. Thesis.

Cahn (1989). *Generating Expression in Synthesized Speech.* Massachusetts Institute of Technology: Master Thesis.

Carlson, R., Sigvardson, T., and Sjölander, A. (2002). Data-driven formant synthesis. *Proc of Fonetik 2002*. TMH-QPSR, 44: 121-124.

Chris, R. (1991). *Speech Processing.* UK: McGraw-Hill, Inc.

Christine, H. S. and Robert, I. D. (2001) Prospects for Articulatory Synthesis: A Position Paper. *4th ISCA Tutorial and Research Workshop on Speech Synthesis.* August 29th - September 1st. Scotland: ISCA2001, 41-44.

Christof, T. (2002). *SVOX: The implementation of A Text To Speech system for German.* Swiss Federal Institute of Technology: Ph.D. Thesis.

Cox, R. V., Kamm, C. A., Rabiner, L. R., Schroeter, J. and Wilpon J. G. (2000). Speech and Language Processing for Next-Millennium Communications Services. *Proceedings of the IEEE*. Vol. 88, No. 8: 1314-1337.

Darwis, H. M. (1987). *Struktur sintaksis ayat selapis dalam Bahasa Malaysia : satu analisis transformasi generatif.* Kuala Lumpur : Dewan Bahasa & Pustaka.

Donovan, R. (1996). *Trainable Speech Synthesis.* Cambridge University: Ph.D. Thesis.

Dusterhoff, K. and Black, A. (1997). Generating F0 contours for speech synthesis using the Tilt intonation theory. *Proceedings of ESCA Workshop of Intonation*, September. Athens, Greece: ESCA, 107-110.

Farid, M. O. (1980). *Aspects of Malay Phonology and Morphology*. Bangi: Universiti Kebangsaan Malaysia.

Fordyce, C. (1998). *Prosody Prediction for Speech Synthesis using Transformational Rule-based Learning*. Boston University: Master Thesis.

Geoff, B. (1984). *Electronic Speech Synthesis.* US: McGraw-Hill, Inc.

Gold, B. and Morgan, N. (2000). *Speech and Audio Signal Processing: Processing and Perception of Speech and Music.* US: John Wiley and Sons, Inc.

Hirst, D., Rilliard, A. and Aubergé, V. (1998). "Comparison of subjective evaluation and an objective evaluation metric for prosody in text-to-speech synthesis." *Third ESCA/COCOSDA Workshop on SPEECH SYNTHESIS*. November 26-29. Jenolan Caves, Blue Mountians, Australia: ESCA 1-4.

Hong J. H. (1997). *Sintesis Ucapan Melayu.* Universiti Kebangsaan Malaysia: Bachelor Thesis.

Jan, Richard, W. S., Joseph, P. O., and Hirschberg, J. (1996). *Progress in Speech Synthesis.* US: Springer-Verlag.

Janet E., and Sangho, L. (1999). *Tree-based Modeling Prosody for Korean TTS System.* Korean Advanced Institute of Science and Technology: Ph.D. Thesis.

Johan, W. (1996). *Analysis and Synthesis Of Degree of Articulation.* Katholieke Universiteit Leuven (KUL): Master Thesis.

John, R. D., John, G. P., John, H. L. H. (1993). *Discrete-Time Processing of Speech Signal.* US: Macmillan, Inc..

Jurafsky, D. and Martin, J. H. (2000). *Speech and Language Processing.* University of Colorado, Boulder: Prentice Hall, Inc.

Kenneth, C. (1997). Stress Assignment in Letter to Sound Rules for Speech Synthesis. *ACL Anthology A Digital Archive of Research Papers in Computational Computational Lingustics.* March, North American 23(1). 246-253

Klatt, D.H. (1987) Review of Text-to-Speech Conversion for English. *Journal of the Acoustical Society of America*. 82(3): 737-793.

Kuek, T. S. (1998). *Sintesis Ucapan Melayu 2 (S.U.M 2) Sistem Teks-ke-Ucapan Melayu Berdasarkan Model KLSYN88.* Universiti Kebangsaan Malaysia: Bachelor Thesis.

Lemmetty S. (1999). *Review of Speech Synthesis Technology.* Helsinki University of Technology: Master Thesis.

Lenzo, K. and Black, A. (2000). Diphone collection and Synthesis. *Proceeding of the International Conference on Speech Language Processing 2000.* October 1-2. Beijing, China: ICSLP2000, 223-237.

Lewis, E. and Mark, T. (1999). Word and Syllable concatenation in Text-To-Speech Synthesis. *In Sixth European Conference on Speech Communications and Technology*. September 1999, Australia: ESCA, 615-618.

Luis, M. T. (1997). *Speech Coding and Synthesis Using Parametric Curves.* University of East Anglia: Master Thesis.

Macon, M., Cronk, A., Wouters, J. and Kain, A. (1997). OGIresLPC: Diphone synthesiser using residual-excited linear prediction. *Proceeding of Department*

of Computer Science Oregon Graduate Institute of Science and Technology. September 1997. CSE-97: 007.

Mangold, H. (2001). Speech Technology in Reality - Applications, Their Challenges and Solutions. *Text, Speech and Dialogue 4th International Conference, TSD 2001*. September 11-13. Zelezna Ruda, Czech Republic: LNAI 2166, 197-201.

Martin, J. (1997). Re-Engineering Letter-to-Sound Rules. *ACL Anthology A Digital Archive of Research Papers in Computational. Computational Lingustics*, March 1997: ACL97.

Möbius B., Sproat R., Santen J. and Olive J. (1997). The Bell Labs German Text-to-Speech System: An Overview. *Proceedings of the European Conference on Speech Communication and Technology*. vol. 5: 2443-2446.

Nik, S. K., Farid, M. O. and Hashim, M. (1989). *Tatabahasa Dewan: Perkataan.* Kuala Lumpur: Dewan Bahasa Dan Pustaka.

Onn, H. M. (1993). *Binaan dan Fungsi Perkataan dalam Bahasa Melayu: Suatu Huraian dari Sudut Tatabahasa Generatif.* Kuala Lumpur: Dewan Bahasa Dan Pustaka.

Pagel, V., Lenzo, K. and Black, A. (1998). Letter to sound rules for accented lexicon compression. *Proceeding of the International Conference on Speech Language Processing 1998.* August 2-4. Syndey, Australia: ICSLP98, 2015-2020.

Parsons, T. W. (1987). *Voice and Speech Processing.* US: McGraw-Hill, Inc.

Rabiner, L. (1993). *Fundamentals of Speech Recognition*. US: Prentice Hall, Inc.

Rabiner, L. R., Jackson, L. B., Schafer, R. W. and Coker, C. H. (1971). A Hardware Realization of a Digital Formant Speech Synthesizer. *IEEE Transaction on Communication Technology*, Vol. COM-I9 (6): 1016-1020.

Raminah, S. and Rahim, S. (1987). *Kajian Bahasa untuk Pelatih Maktab Perguruan*. 8th ed. Petaling Jaya: Penerbit Fajar Bakti Sdn. Bhd.

Rilliard, A. and Aubergé, V. (2001). Prosody evaluation as a diagnostic process: subjective vs. objective measurements. *4th Speech Synthesis Workshop*. August 29 - September 1. Scotland, ISCA140-144

Ronald, A. C. (1995). *Survey of the State of the Art in Human Language Technology.* National Science Foundation, Oregon Graduate Institute.

Rowden, C. (1992). *Speech Processing.* UK: McGraw-Hill, Inc.

Shuzo, S. and Kazuo, N. (1985). *Fundamentals of Speech Signal Processing.* UK: Academic Press Japan, Inc.

Sheikh, H. S. S. (1993). *A Comparative Study of the Traditional Classifier and the Connectionist Model for Speaker Dependant Speech Recognition System.* Universiti Teknologi Malaysia: Master Thesis.

Syrdal. A., Bennett, R and Greenspan, S. (1994). *Applied Speech Technology.* UK: CRC Press, Inc.

Syrdal, A., Moehler, G., Dusterhoff, K., Conkie, A and Black, A. (1998). Three Methods of Intonation Modeling. *3rd ESCA Workshop on Speech Synthesis*. November 26-29. Jenolan Caves, Australia, 305-310.

Tan, T. S., Sheikh, H. and Aini, H. (2003). Building Malay Diphone Database for Malay Text to Speech Synthesis System Using Festival Speech Synthesis System. *Proc of The International Conference on Robotics, Vision, Information and Signal Processing 2003*. January 22-24. Penang, Malaysia: ROVISP03, 634-648.

Tan, T. S. and Sheikh H. (2003). Building Malay TTS Using Festival Speech Synthesis System. *Conference of The Malaysia Science and Technology*, September 2-3. Johor Bahru, Malaysia: MSTC 2002, 120.

Taylor, P., Black, A. and Caley, R. (1998). The architecture of the Festival Speech Synthesis System. *3rd ESCA Workshop on Speech Synthesis*. November 26-29. Jenolan Caves, Australia. ESCA: 147-151

Taylor, P., Black, A. and Caley, R. (1999). Festival Speech Synthesis System: system documentation (1.4.0). *Human Communication Research Centre Technical Report.* HCRC/TR, 83-202.

Thierry, D. (1993). *High QualityText-To-Speech Synthesis of the French Language.* Faculté Polytechnique de Mons: Ph.D. Thesis.

Ting, H. N. (2002). *Speech Analysis and Classification using Neural Networks for Computer-based Malay Speech Therapy.* Universiti Teknologi Malaysia: Master Thesis.

Witten, I. H. (1982). *Principles of Computer Speech.* US: Academic Press, Inc.

Yarrington, D., Bunnell, H.T., and Ball, G. (1995). Robust Automatic Extraction of Diphones with Variable Boundaries. *Proceedings of the 4th European Conference on Speech Communication and Technology*: 1845-1848.

Yeoh, P. Y. (2002). *Text-to-speech for Malay language.* Universiti Teknologi Malaysia: Bachelor Thesis.