TEXT-TO-SPEECH CONVERSION (FOR BAHASA MELAYU)

By

MUHAMMAD ASHRAFF BIN AHMAD

Dissertation submitted in partial fulfillment of
the requirements for the
Bachelor of Technology (Hons)
(Information Technology)

NOVEMBER 2004

Universiti Teknologi PETRONAS
Bandar Seri Iskandar
31750 Tronoh
Perak Darul Ridzwan

CERTIFICATION OF APPROVAL

Text-to-Speech Conversion (for Bahasa Melayu)

by

Muhammad Ashraff bin Ahmad

A project dissertation submitted to the
Information Technology Programme
Universiti Teknologi PETRONAS
in partial fulfillment of the requirement for the
BACHELOR OF TECHNOLOGY (Hons)
(INFORMATION TECHNOLGY)

Approved by,

____

(Mr. Jale Ahmad)

UNIVERSITI TEKNOLOGI PETRONAS
TRONOH, PERAK
November 2004

CERTIFICATION OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this project, that the original work is my own except as specified in the references and acknowledgements, and that the original work contain herein have not been undertaken or done by unspecified sources of persons.

MUHAMMAD ASHRAFF AHMAD

ABSTRACT

Text-to-Speech (TTS) is an application that help user in having the text given to be read out loud. This project highlighted in creating a TTS system that allows text reading in Standard Malay Language (Bahasa Melayu). There is a lack of computer aided learning (CAL) tools that emphasize in Malay linguistic and misconception that people have regarding the usage of English-based TTS to read Bahasa Melayu text derived the development of this project. The end result is the TTS conversion prototype for Bahasa Melayu that reads by syllable using syllabification techniques through the employment of Maximum Onset Principle (MOP) and produce syllable sounding speech by using syllable to sound mapping method.

Keyword: Text-to-Speech, syllabification, maximum onset principle, syllable to sound mapping.

ACKNOWLEDGMENT

I would like to express my utmost sincere gratitude to the university and following people who had contributed and given tremendous effect to me throughout the development of this project.

Mr. Jale Ahmad, my supervisor for his willingness to take me under his supervision and giving me the freedom to do this on my way. His ideas, guidance and support had given me the courage to complete this project and to overcome many problems that I had encountered especially in the beginning of the semester where I had almost hit the dead end. Thank you for giving me my own wings and decide on what I want to do and not to solely depend on others in completing and making decision.

My special credit to Ms. Norshuhani Zamin for her concern and accompany that helped me so much in putting myself together and look up straight on what should I be doing and look for throughout the accomplishment of this project.

My beloved parent who had inspired me in completing this challenge and makes it through no matter how hard it might seems to be. Thank you for the love and space given to me. Not to forget, lecturers who had given me plenty of ideas and advices throughout this semester. My deepest gratitude goes to Mr. Suhaimi Abdul Rahman, Mr. Faizal Ahmad Fadzil and Ms. Hasiah Mohamed.

Finally, this project would not be completed without acknowledging the contributions of my friends, for being very helpful and supportive throughout the development of this project. There are too many names to mention and to list some and letting some in blank would not be appropriate. Thus, I would like to personally thank every single of them.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

There are over 500 million people who speak Bahasa Melayu in the developing southeastern nation of Malaysia, Indonesia, Brunei, Singapore and Southern Thailand (Yousif A. El-Imam, Zuraida Mohammed Don, 2000).

With such a wide coverage of people using this language, it shows that Bahasa Melayu is vulnerable and competitive as other major languages in this world such as French, Japanese and English to be learned and used for knowledge exchange and communication medium.

Hence, I believe that the development of TTS for this language is essential in becoming one of the tools in revolutionizing this language into a bigger scope of user. This work examine the nature of TTS and its important component such as syllabification and sound synthesizing and how will it be done and implement in producing a TTS application.

Apart from that, it will also include theories, rules and boundaries employed for the development of this application. Those who have basic knowledge on Bahasa Melayu phonology and orthography may have a slight advantage in understanding further content of the dissertation.

1.1 Background of Study

A Text-To-Speech (TTS) synthesizer is a computer-based system that should be able to read any text aloud and focused on the automatic sound production of new sentences. Different from Voice Response Systems that simply concatenate isolated words or parts of sentences from stored vocabularies, TTS is referred as the automatic production of speech, through a grapheme-to-phoneme transcription of the sentences to utter (Thierry Dutoit, 1996).

The objective of TTS system is to be able to read and extract given text and produce sounding speech as natural as possible. TTS conversion problem is a problem that spans the domain of acoustics, digital signal processing, and natural language processing (Yousif A. El-Imam, Zuraida Mohammed Don, 2000).

Thus, good choice of synthesis performed through phonological analysis of the target language is needed to obtain good quality natural speech. However, this project did not focus on creating natural speech but producing almost natural sounding speech whereby syllable sounding is concerned.

In general, TTS is composed of two parts – the front end and the back end. The front end takes input in the form of text and output a symbolic linguistic representation. The back end takes the symbolic linguistic representation as inputs and outputs the synthesized speech waveform. The naturalness of a speech synthesizer usually refers to how much the output sounds like the speech of a real person.

Although Bahasa Melayu used similar alphabetic and almost similar orthographic system as English in the written form of the language, both differ in terms of phonetics and linguistic property of the language. For example the word "padi" (paddy) will carry different syllable and phonetic structure from English.

"padi" will be syllabified as <pa – di> and pronounced as /padi/ in Bahasa Melayu but in English, it is pronounced as /pædaɪ/ although they shared the same syllable representation. This clearly distinguished that using English based TTS in reading

Bahasa Melayu words or sentences would not exactly follow correct pronunciation and articulation. However, there is a slight similarity in the orthography (syllabification and syllable structure) of the word used in both languages. This will be later described in upcoming chapter.

1.2 Problem Statement

It is very rare to find a TTS system that converts Bahasa Melayu text into speech. The fact that most TTS engines mainly developed and tested for profitable language such as English, French, German, Spanish, Japanese and Chinese (Ksenia Shalonova, Roger Tucker, 2003), had initiated the development of this project.

Moreover, it is undeniable that there is still lack of computer aided learning (CAL) tools that emphasize in Malay linguistic that help students especially the non-mother tongue speakers in learning the language effectively.

Furthermore, the thought of using English-based TTS converter to read Bahasa Melayu text has to change due to its dissimilarities of orthography and phonological representation.

The fact that English-based TTS uses the same alphabetic system as Bahasa Melayu, most user thought that they can use English-based TTS converter to read Bahasa Melayu text. However, it does not work that way.

The significant of this project is to open up new horizon in TTS technology whereby it can provide alternatives to the current existed TTS application. With the existence of Bahasa Melayu version of TTS application, it is hope that it will expand the Computer Linguistic horizon especially in the Malay linguistic development.

1.3 Objectives

There are two main objectives for this project. The main objective of this project is to create a TTS application that reads Bahasa Melayu texts aloud. Apparently, there is still little effort was put in this area where most development was more focused on popular languages. With the development of this project, we hope that it can create a firm foundation in local computer linguistic field and promoting the national language into a bigger market in future time.

The second objective is to illustrate the method of producing speech sound through syllabification techniques and syllable to sound mapping system. This is to show how the end results of the prototype is achieved and what are the steps taken in producing the output.

1.4 Scope of Project

This project will focus on creating TTS conversion method and prototype for Bahasa Melayu that reads by syllable sounding using syllabification techniques and syllable to sound mapping method. The target is to achieve a minimum of one word to be read and syllabified correctly according to several rules and boundaries that is further described in coming chapters.

The input to the system at present is only a piece of Bahasa Melayu text and will not process a text in batch or a text file. The input can have a minimum of monosyllabic words but will not be able to process words with syllabic consonants.

In addition, this system will not cater for loanwords, punctuations, abbreviations, acronym and numbers (for examples, please refer to APPENDIX C). The input for this project will need to have a maximum of 2 consonant clusters in a given word regardless of it preposition within the orthographic structure and follow by given rules to ensure smooth text to sound processing. Since this project is a prototype, it will only produce up to certain capacity of result. Further details will be later described throughout the report.

Within approximately 5 months of total development and researches, the target of this to achieve a functional and working product that at least will be able to read minimum of one word correctly. Scheduled tasks for total development of this project are summarized in Gantt chart attached in APPENDIX A.

CHAPTER 2

LITERATURE REVIEW

In this literature review, the goal is to learn on what other researchers have done or contributed to the discipline that I am working on, to help gain more relevant materials and support for my project.

In general, Text-to-Speech in layman term refers to the conversion of normal text into audible speech. This presumption was supported by WIKIPEDIA, The Free Encyclopedia from the internet that defines Text-to-speech or Speech synthesis as "The artificial production of human speech. A system used for this purpose is termed a speech synthesizer, and can be implemented in software or hardware. Speech synthesis systems are often called text-to-speech (TTS) systems in reference to their ability to convert text into speech. However, there exist systems that can only render symbolic linguistic representations like phonetic transcription into speech".

Similar definition stated that TTS as the automatic production of speech, through a grapheme-to-phoneme transcription of the sentences to utter. It is said that a Text-To-Speech (TTS) synthesizer is a computer-based system that should be able to read any text aloud, whether it was directly introduced in the computer by an operator or scanned and submitted to an Optical Character Recognition (OCR) system (Thierry Dutoit, 1996).

Conclusively, TTS can be simplified as a system that read any text, extract useful information from it and use it to compose as naturally sounding speech as possible (Yousif A. El-Imam and Zuraida Mohammed Don, 2000).

TTS system uses many approach, most popular approaches is rule based and dictionary based. The dictionary based approach basically stored certain capacity of phonological knowledge into list of words or lexicon in a database for synthesis reference. This approach normally consumes quite a reasonable size due to its phonological representation of lexicons (Thierry Dutoit, 1996).

This concept has been adopted by the Japanese TTS system where morphological analysis and dependency analysis relation employing semantic categories techniques is used to store up to 430 000 words from dictionary to the system (Yoshifumi Ooyama, Masahiro Miyazaki, Satoru Ikehara, 1987).

On the other hand, the rule based system transfer most of the phonological competence of dictionary into set of letter-to-sound or grapheme-to-phoneme rules. By employing certain set of rules, the grapheme-to-phoneme component will determine the phonetization of the sentences (Yousif A. El-Imam, Zuraida Mohammed Don, 2000). These rules will be an input orthography to create the appropriate sound of the language.

This system however requires less storage and produces a set of intonation events where the characteristics of the syllable later produces sound (Joram Meron, 2001). Though both methods had been fairly argued on its reliability accuracy, both methods seem to have its own advantages and disadvantages (Thierry Dutoit, 1996).

Generally, TTS is composed of two parts – a front end and a back end. The front end takes input in the form of text and output a symbolic linguistic representation. The back end takes the outcomes as inputs and outputs the synthesized speech waveform. The naturalness of a speech synthesizer usually refers to how much the output sounds like the speech of a real person.

The front end has two major tasks. First, it takes the raw text and converts into their written-out word equivalents. This process is often called text normalization, pre-processing, or tokenization (Roxanne Ruzic and Kathy O'Connell, CAST). The goal is to convert input text into phonological description consisting in a phoneme chain associated to some sort of prosodic and accentual description (Romain Prudon, Christophe d'Alessandro, 2002).

The process of assigning phonetic transcriptions to words is called text-to-phoneme (TTP) or grapheme-to-phoneme (GTP) conversion. The combination of phonetic transcriptions and information about prosodic units make up the symbolic linguistic representation output of the front end.

However there are other methods apart from phonemic representation such as the syllabic representation (Yousif A. El-Imam, Zuraida Mohammed Don, 2000). This method uses the GTP or TTP conversion from normalized text and groups the phones into syllables and sound creation will be based on the syllable extraction made (Paul C. Bagshow, Briony J. Williams, 1992). Syllabification method is used in the development of this project.

Syllable varies in weight or complexity. It is pattern in many languages in ways that support or enable prosodic structuring in speech which can be use as speech synthesizing and recognizing method (Li Zhang, William H Edmondson, 2002).

Three types of syllable extraction rules were established; those proposed by phonologist, those that people apply intuitively in metalinguistic tasks and those that listener may use during online lexical access (Caroline Floccia, Jeremy Goslin, Naanaa Bouketir, Joel Bradmetz, 1999).

Technically, syllable is referred as a timing unit for language. Words in language take certain amounts of time to utter, where the time measurement is associated with syllable usage. The structure of the syllable can be described or analyzed in various ways but the most obviously, it is viewed as a group of segments that contains a vowel nucleus with up to three consonant before and after (William H Edmonson and Li Zhang, 2002).

There are quite a few techniques or rules in syllabifying a text. Some of these rules were specially catered for certain language such as French but there are rules that were meant to be used for any languages.

The Kahn's algorithm or Maximum Onset Principle (MOP) is one of the rules that can generally be used in most languages. According to the principle, consonantal phones were assigned to one syllable onset to the maximum number as possible as long as the resulting cluster can be found at the beginning word in the language (Caroline Floccia, Jeremy Goslin, Naanaa Bouketir, Joel Bradmetz, 1999).

William H Edmonson and Li Zhang outlined a model to symbolize the syllable structure that compliment with the MOP principle. The model described syllable to be decomposed into two entities; the onset and rhyme. Rhyme on the other hand will be segmentized into two categories, the nucleus and coda. This structural approach will work up close with the MOP to work with phonotactic constrains to identify syllables in sequence (William H Edmonson and Li Zhang, 2002).

Looking on the other side of the TTS cycle; the back end, it takes the symbolic linguistic representation derived from the GTP or syllabified phoneme and converts it into actual sound output. The back end is often referred to as the synthesizer (Roxanne Ruzic and Kathy O'Connell, CAST).

The synthesizing process in sound production is highlighted as the synthesis unit generation. It is where the output of GTP or syllabified phoneme is used as input to this module.

For the development of this project, I favor generating the synthesis unit from the syllable due to fact that considering the constraint given had fairly basic phonemic representation and prosodic features (Yousif A. El-Imam and Zuraida Mohammed Don, 2000).

In producing the audible sound derived from the syllables, I opted for a syllable to sound mapping where each syllable extracted from a word will be directly map to most accurate sound or best matching sequence through a set of lookup table stored in a database.

This is quite similar to unit selection method. The unit selection process is based on some type of dynamic programming that selects from the speech database unit with minimal cost functions (T. Lambert, A.P. Breen, B. Eggleton, S.J. Cox, B.P. Milner, 2003). The selection is performed after the text processing module and then decided on where events should take place and creates a vector for these events before the selection runs through the database finding best matching sequence (Joram Meron, 2001).

The main motivation of using this method is to create a more natural sounding speech and easier mapping compared to other method of synthesizing. This method consume large speech database, unit selection synthesis can take care of variability in speech without complex signal and linguistic processing needed to model phonetic variations and speech prosody (Yousif A. El-Imam and Zuraida Mohammed Don, 2000).

Therefore, by using this concept, another approach that uses the basic concept of unit selection is constructed. The construction of the new method employed the usage of pre-stored speech sound in a database to create speech prosody in a form of syllable sounding speech output.

As most system developed are more concentrated to more popular and profitable language (Ksenia Shalonova, Roger Tucker, 2003), it is very rare to find a TTS system that converts Bahasa Melayu text into speech. Thus, through this project, I wish to introduce a TTS system for Bahasa Melayu.

In this paper, I will present a method of producing a text to speech converter that uses syllabification technique with the help of Maximum Onset Principle to create syllable and identify syllable boundaries and produce audible sound through syllable to sound mapping method. As a result, the application developed should be able to transform written text into speech sound. The methodology and implementation are discussed in following chapters.

CHAPTER 3

METHODOLOGY/ PROJECT WORK

3.1 Procedure Identification

Throughout the project completion, this project implies two main activities. The first activity involved is information mining and theory construction on text-to-speech technology and its architecture. Secondly, the development of prototype with the use of programming tools and audio processing tools identified.

In developing the application, the System Development Life Cycle is used to guide the project throughout its development. The life cycle is divided into six stages:

- System initiation and feasibility study
- Project planning and functional analysis
- System Design
- Programming
- Testing and implementation
- Post-implementation evaluation/ maintenance

3.2 Project Work

3.2.1 Overview of Bahasa Melayu Spelling

Since the ancient times, Indian and Chinese records have referred to Malay language playing the role of a lingua franca for the Malay Archipelago region. The language continued to improve in the 20th century throughout the colonial period and with the concerted efforts of Dewan Bahasa dan Pustaka, the Malaysia's language planning agency (Nik Safiah Karim, 1995).

Phonology and orthography are branches of linguistics and important components for connecting letter-to-sound or sound-to-letter correspondence. Many definitions are associated to these terms but similarly they were focused on the study of letters and sounds for a particular language.

Macmillan English Dictionary (2002) define each term as "the study of the pattern of speech sounds used in a particular language" and "the system of spelling that a language uses". It can be concluded that phonology is a study of the sounds of the speaker's language and orthography can be concluded as a system that uses written symbols to uniquely identify each other through the representation of alphabet and graphemes.

There are 26 letters used to create graphemes. Letters in alphabet are the raw material used to create grapheme which in return creating phonemics representation of the language. Dewan Bahasa dan Pustaka outlined Malay graphemes that consist of single letter and combination of letters as following:

- **Single vowel letters:** a e i o u (e has two different sound denoted by e-taling and e-pepet)
- **Single consonant letters:** b c d f g h j k l m n p q r s t v w x y z

- **Letter combination:** a sequence of two or more different letters used to represent single sound or sound sequence.
    - **Diphthong:** combination of two vowels and pronounce as one sound. Examples of diphthong are ai, au and oi.
    - **Consonant Cluster (diagraph):** there are five valid consonant clusters that produced single sound consonant. Examples of these clusters are gh, kh, ng, ny, and sy.

Vowels and consonants are major entity in speech sounds. The vowels are syllabic while consonants are non-syllabic. Moreover, consonant are sounds produced with stricture and the position of the organ that produces the sounds can be identified precisely while vowels are not. Through these characterization of graphemes distinguished the distribution of sound and production of sound.

Every speech sound is produced by articulation or the movement of one or more vocal organ along the vocal tract. When articulation comes into count, it is inevitably to discuss on place of articulation and manner of articulation. Table 3.0 and 3.1 illustrate the place of articulation and manner of articulation for consonants and vowel in Bahasa Melayu taken from Yousif A. El-Imam and Zuraida Mohammed Don (2000).

Place and manner of articulation	Bilabial	Labiodentals	Dental	Alveolar	Post-alveolar	Palato-alveolar	Palatal	Velar	Uvular	Glottal
Oral stop	/p/ /b/			/t/ /d/				/k/ /g/		
Nasal (stop)	/m			/n/						
Affricate						/ts/ /dz/				
Fricative		/f/ /v/		/s/ /z/					/x/	
Lateral				/l/						
Approximant					/r/		/j/	/w/	/h/	

Table 3.0 Articulation for consonants of Bahasa Melayu

Tongue position/ height	Front	Central	Back
High	i		u
Mid-high	e		o
Mid		δ	
Mid-low		a	

Table 3.1 Articulation for vowels of Bahasa Melayu

To describe the place of articulation of individual consonants and vowel, the vocal tract is divided into different segments or regions. The locations of these organs and regions are as shown in the following table:

Vocal organs and regions of articulator	**Adjectives**
Nose	Nasal
Mouth	Oral
Lips	Labial
Teeth	Dental
Alveoli (or alveola ridge or gum ridge)	Alveolar
(hard) Palate	Palatal
Velum (soft palate)	Velar
Pharynx	Pharyngeal
Uvula	Uvular
Larynx	Laryngeal
Glottis	Glottal

Table 3.2 Name of vocal organs and region of articulator

In accordance to the table above is the diagram that illustrates the articulators used frequently in the study of phonetics.

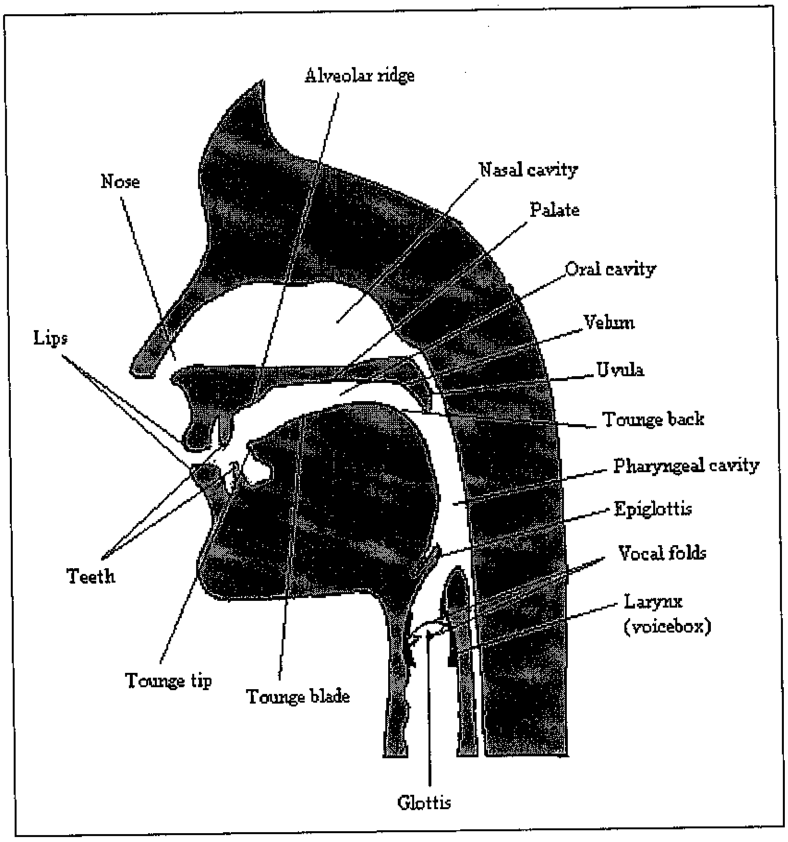


Figure 3.0 Diagram of vocal organs and articulary regions

There are many things that can be discussed regarding consonants and vowels such as its characteristics, variations and system. However, since my study is concentrated into syllabification and its sound, I shall concern to this area of research. Further details regarding this matter should best referred to phonology and phonetic texts and references.

3.2.2 Syllable Structure

In layman term, syllable is best described as an important unit of spelling and acts as a timing unit for language. Most speaker of Bahasa Melayu would have no problem in cutting up words into syllable since syllable is countable. Bahasa Melayu speaker will be able to identify the word "makanan" (food) carry more than one syllable but they might have different view of segmenting them.

Generally, a syllable consists of consonant, followed by a vowel (V) and followed by another consonant (C). In Bahasa Melayu, a syllable may also contain only vowel as outlined in the four common syllable pattern system. Conventionally, the syllable will follow the following syllable structure model shown in Figure 3.1

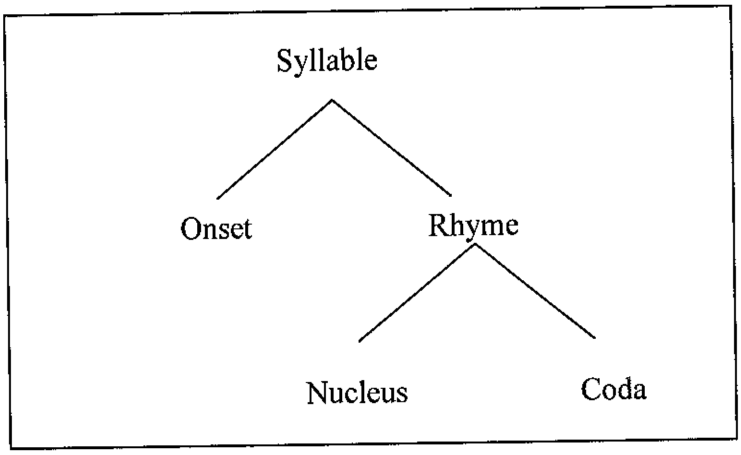


Figure 3.1 Conventional Syllable Structure

From this figure, most researchers agree that nucleus is the basic element in the syllable structure. This is supported by the work of William H Edmonson and Li Zhang (2002) that describe their syllable structure as "group of segment" consists of a vowel (nucleus) and with up to three consonant following the vowels (onset and coda). Their notion can be described as (C) (C) (C) V (C) (C) (C) where brackets show the characters presence in respective position.

However, in Bahasa Melayu, though they followed the same conventional syllable structure, the maximum number of consonant that can exist in a syllable is three consonants assigned to onset and two consonants assigned to coda respectively. This can be denoted by (C) (C) (C) V (C) (C).

Though such condition occurs, due to the limitation set for this project, the study will cover up to the maximum of two consonant to be assigned respectively to onset and coda. The reason behind this is due to the fact that triple consonant in Bahasa Melayu syllable will represent loanword.

For example the word "struktur" came from English word "structure" and the word "skrip" came from "script" from English. The usage of loanwords in this study is also part of limitation discussed in previous chapter.

In addition, William H Edmonson and Li Zhang (2002) also add the principle of the Sonority Sequencing Principle into their structure to give the syllables sonority value. Sonority is defined by the opening of the air passages; the more it opens the more sonorous the element is.

This principle makes nucleus as the most sonorous element and the sonority wave build up from less sonorous element i.e. onset to the nucleus (maximum sonority) and progressively decrease when passing coda.

Following example is replicated from William and Zhang's principle and applied in context of Bahasa Melayu syllable structure.

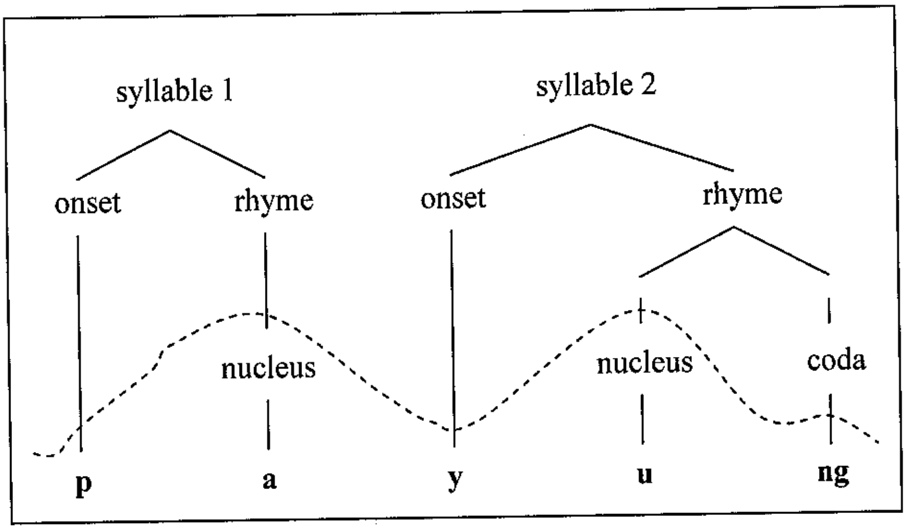


Figure 3.2 Example of William and Zhang's approach on syllable structure for Bahasa Melayu

*Note: The word "payung" carry the meaning of umbrella in English

As mentioned above, the smallest unit to create a syllable in Bahasa Melayu is a single vowel itself. This referred as not having coda in the syllable structure. By examining the sequence of the consonant and vowel would give a brief understanding pertaining permissible syllable structure in Bahasa Melayu.

Taken from Dewan Bahasa dan Pusataka, Bahasa Melayu has four basic syllable pattern and six extended syllable pattern. Following are samples of words and its syllable pattern.

Syllable Pattern	**Example**
V	"*a*-nak" (child)
VC	"*om*-bak" (wave)
CV	"*to*-lak" (push)
CVC	"*pin*-tu" (door)

Table 3.3 Basic syllable pattern

Syllable Pattern	Example
CCV	"*sya*-ri-kat" (company)
CCVC	"*trak*-tor" (tractor)
VCC	"*emm*" (type of human expression)
CVCC	"*peng*-san" (faint)
CCVCC	"kom-*pleks*" (complex)
CCCV	"*stra*-te-gi" (strategy)
CCCVC	"*struk*-tur" (structure)

Table 3.4 Extended syllable pattern

From the tables above, we can identify that most complex syllable pattern occurs when there is loanwords involved in the syllabification. Therefore, to meet with my objective and scope of project, I had made several assumptions pertaining to this matter.

1. Since most loanwords especially taken from English have three consonant clusters before vowel (CCCV, CCCVC), any pattern that follow triple consonant cluster will be consider void.

2. It is also impossible for a normal standard Bahasa Melayu to have vowel in between double consonant (CCVCC) or double consonant being position as onset and single consonant being as coda (CCVC) since it will only occur to loanwords.

3. Since this project caters for standard Bahasa Melayu, colloquial words and expressions is not within my scope of study. Thus, syllable containing vowel followed by double consonant (VCC) will consider void to this system.

In this study, the speech production will follow the syllable structure made from lexicon syllabification. Hence, it is important to understand the characteristics and its pattern in order to come out with best syllabification technique and accurate word segmentation approach.

To perform syllabification for this project, I had chosen the Maximum Onset Principle (MOP) as my approach. This principle will be implemented after the preprocessing of the text is performed within the process flow described in the next sub-chapter.

3.3 Methods and Approaches

In this study, there are two main approaches that I use in conducting the project. The first approach employed is the Maximum Onset Principle that will handle the word division into sets of syllables and syllable-to-sound mapping method. Each approach occurs in the beginning and towards the end of process flow respectively.

3.3.1 Overview of Process Flow

In performing the text-to-speech process cycle, a module is developed to show the process flow that happen throughout the project development. Figure 3.3 illustrate the process flow of the TTS system.

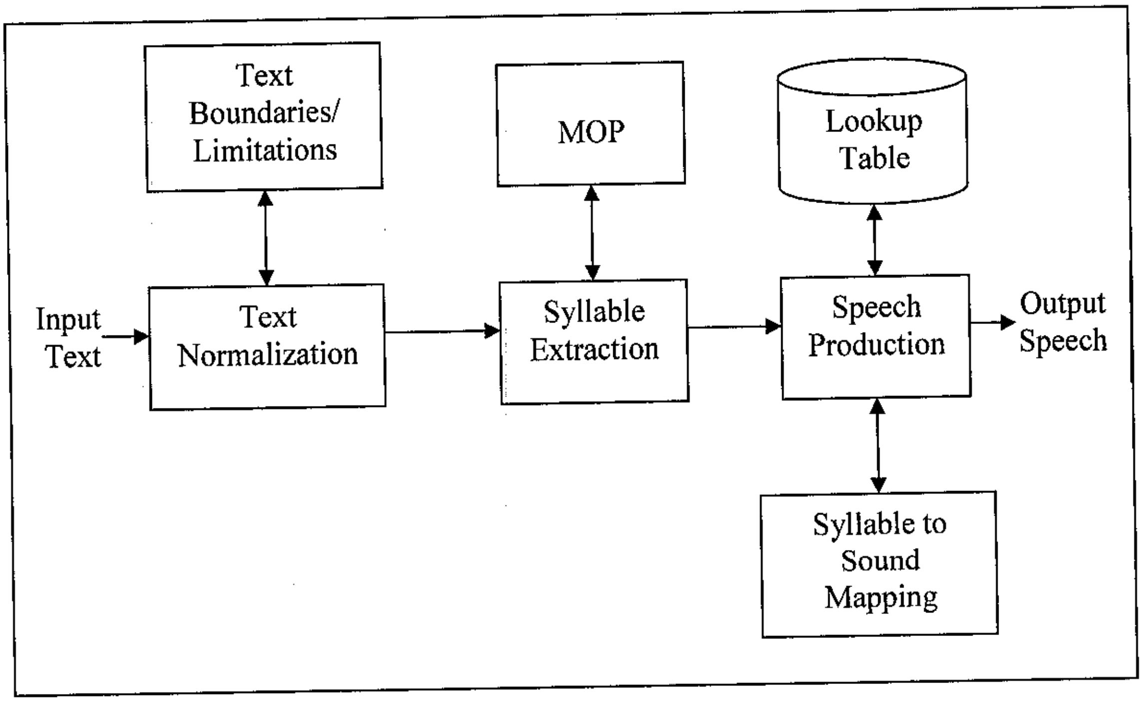


Figure 3.3 Text-to-speech process flows

There are three stages where TTS conversion is concern. The text normalization will perform text preprocessing where it is needed to prepare the input for further processing and analysis by the remaining module of the system. These texts will be referred to the boundaries and limitation that had been set before hand to eliminate unwanted results.

The normalized text will be the input for syllable extraction module where Maximum Onset Principle will be applied on the processed text to create a syllable boundary for the text and produce syllable structure before it will be mapped to the sound.

Finally, speech production is performed from the previous module's output. Through this module, a syllable will be mapped to its most matching sequence of speech audio from a lookup table stored in a database. The output from this process will be a syllable sounding speech. We will look in further description of each process in next sub-heading of this chapter.

3.3.2 Text Normalization

Text normalization is also referred to tokenizing or preprocessing. This stage will revolve around accepting the right combination of words given to the system and prepare it for the execution of next process stage.

As stated in previous chapter, this system will not cater for punctuations, numbers, acronyms and abbreviations. Thus, this will be the guideline for the text normalizing to analyze each word entered to the system.

For example, if a text entered is "12" to indicate the word twelve, the system will automatically void this representation and ask for reinsertion of new input. Same case goes to any word that contains other limitation stated above, for example "UTP" as Universiti Teknologi Petronas or "Pn." for "puan" (madam). More examples can be obtained from APPENDIX C.

Apart from that, tokenizing will also look for the orthography of the text entered. In my early description, I had stated that the system will not entertain syllabic consonant. Therefore, an input without a vowel is beyond the system processing capability.

User might enter meaningless text such as "trmplkk" which has no meaning in Bahasa Melayu. Since the basic structure for a meaningful word in Bahasa Melayu must contain vowel between permitted consonant sequences, syllabic consonant is beyond the system capability.

Though it is also stated that the system do not entertain loanword, there might be some that will slip through this stage since some loanwords happened to use simpler syllable pattern especially when it involve with the usage of diphthongs.

Using the assumptions made, identified consonant clusters in a word would possibly determine whether the word entered is a loanword or not. Syllable pattern that follows the sequence of CCCV, CCCVC, CCVCC and CCVC would be identified as a loanword. In the case of diphthong usage in a word, it is barely inevitable for the system to recognize the difference of loanwords to the genuine.

This is because diphthong is assumed as one vowel in a syllable sequence. That would indicate all diphthong (ai,au and oi) as V not VV thus belong to the permissible syllable pattern employed for this system that is the CV pattern. Sample word such as "boikot" (boycott) and "audio" (audio) are few examples that the system might missed throughout this stage.

From all conditions stated above, the input will be validated accordingly and valid input will initiate the next stage of the TTS conversion system which is the syllabification process. Invalid input will be rejected from the system and new value will be asked to begin new process initiation. The overall process flow for this stage is best described by the following illustration:

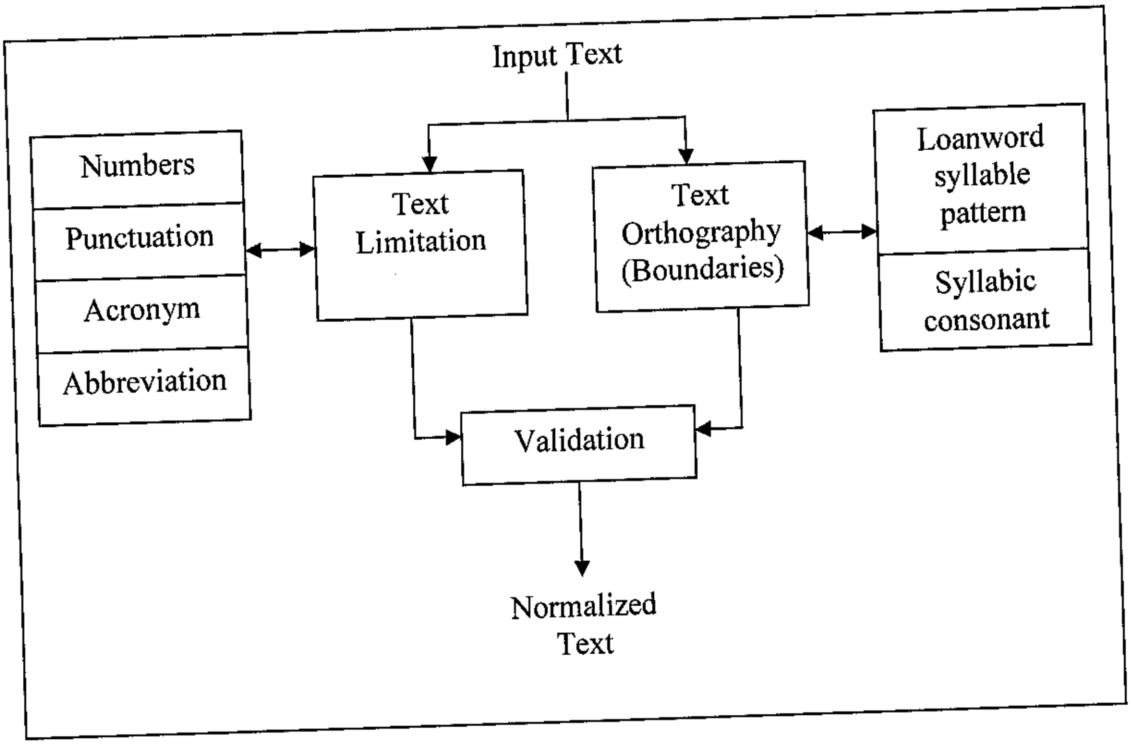


Figure 3.4 Text normalization process flow

3.3.3 Syllable Extraction

In this stage, processed text will be syllabified based on the syllable rules predefined and the use of MOP. In the end of this process, each syllable extracted will be inserted with a syllable boundary to mark the number of syllable exist in the word given. This boundary will be represented by a hyphen ("-") symbol.

To help towards better understanding of how the syllabification process work, one example is given to illustrate the general process of extracting syllable from a given text as follows:

Input word: saudara (relative)

Processing Stage	**Output**	**Note**
1	Sau-dara	In syllabification rules, *au* is a diphthong thus it is consider as a single vowel. Hyphen is inserted after first vowel to indicate first syllable following the rule that division is performed if there is a consonant in a middle of two vowels.
2	Sau-da-ra	The next vowel found indicate another syllable and hyphen is inserted after second vowel to indicate second syllable following the rule that division is performed if there is a consonant in a middle of two vowels.
3	Sau-da-ra	Final vowel is found indicating the final syllable. No hyphen is inserted to show final syllable is found from the word entered.

Table 3.5 Example of Text → Syllable process stage

The number of processing stage differs from one input to another. Longer word or text would require more processing stages. Each processing stage will undergo sets of predefined rule and matching rule will indicate an output (syllable) before the next processes began.

In this context of study, Dewan Bahasa Dan Pustaka (2004) outlined three rules in extracting syllable from a word. The rules are:

1. If there is two vowels subsequently position one after another, syllable is extracted between the vowel clusters. For example:

Word	Syllabified result
main "play"	ma - in
buah "fruit"	bu -ah

Table 3.6 Rule no. 1

However, exception is made on diphthong since both vowels were consider as one. For example:

Word	Syllabified result
ghairah "passion"	ghai- rah
saudara "relative"	sau - da - ra

Table 3.7 Rule no. 1 (case: diphthong)

2. If there is a consonant in the middle of a word, including valid consonant clusters in between two vowels, word division will be performed before the consonant or the consonant clusters. Examples are:

Word	Syllabified result
sangat "very"	sa - ngat
sulit "confidential"	su - lit

Table 3.8 Rule no. 2

3. If there is a non-valid consonant cluster in the middle of a word, division will be performed between the consonant cluster itself. For example:

Word	Syllabified result
mandi "bath"	man - di
mutlak "absolute"	mut - lak

Table 3.9 Rule no. 3

From the rules made and scope that this project is handling, I had summarized a list of permissible sequence that would contribute into word syllabifying. The permissible sequences are diphthong rule, onset sequence and consonant cluster. Following are examples of words used with these sequences.

Vocal cluster	Example		
	Front	Middle	End
ai	*Air* "water"	Gh*ai*rah "passion"	Pand*ai* "cleaver"
au	*Au*dio* "audio"	S*au*dara "relatives"	Harim*au* "tiger"
oi	*Oi*** "Malay expression"*	B*oi*kot* "boycott"	Amb*oi*** "Malay expression"

Table 3.10 Diphthong rule

Note:

* are loanwords

** are emotional expression

Cluster	Sample word
ng	*Ng*antuk "sleepy"
ny	*Ny*anyi "sing"
sy	*Sy*arikat "company"
gh	*Gh*airah "passion"
kh	*Kh*usus "special"

Table 3.11 Permissible onset sequence

Cluster	Sample word	
	Middle	**End (Coda)**
ng	A*ng*in "wind"	Peni*ng* "dizzy"
ny	Ha*ny*a "only"	-
sy	Da*sy*at "terrible"	-
gh	-	Mubali*gh** "missionary"
kh	A*kh*ir "final"	Tari*kh** "date"

Table 3.12 Permissible consonant cluster

Note:

* are loanwords

From the rules set, the next step is to determine the ideal method to separate polysyllabic words into syllables. For this project, I had chosen the Maximum Onset Principle (MOP) as the method to perform syllabification and find the syllable boundaries.

Basically, what MOP does is to syllabify the word according to the vowel existed in a word. The first step in MOP is to identify a vowel since it is the nucleus in a syllable. Next, it will look at the consonant on its left and check with the permissible sequence of consonant cluster with rules identified earlier.

If it conforms to the rules, a symbol is inserted to mark the boundary. Then it will look on the next vowel and perform the same step taken as before. Special procedure is performed when non permissible consonant cluster is found after the first syllable.

In this case, the first consonant will be assigned to the first syllable and the second consonant will belong to the next syllable. Towards the end of the word, the remaining consonants will be assigned to the next identified vowel as its coda clusters.

It can be concluded that MOP prefers onset more than coda. Therefore, constraints are catered for onset only and not coda as the case of invalid consonant cluster explains before. The process of MOP can be illustrated in the following diagrams:

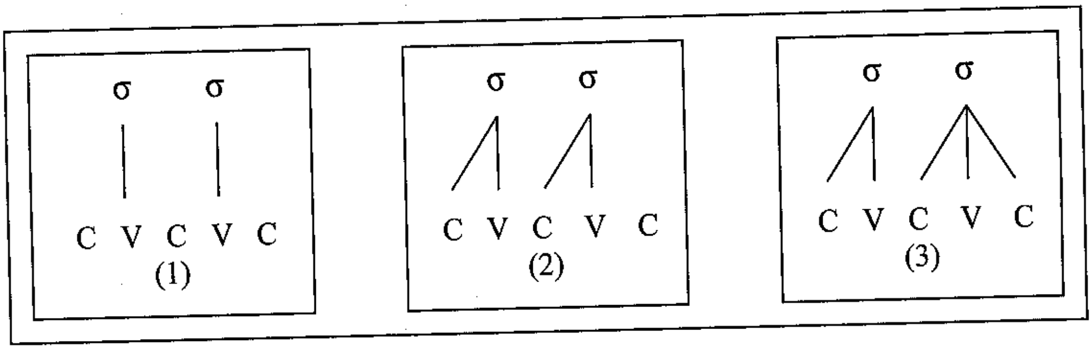


Figure 3.5 Maximum Onset Principle process

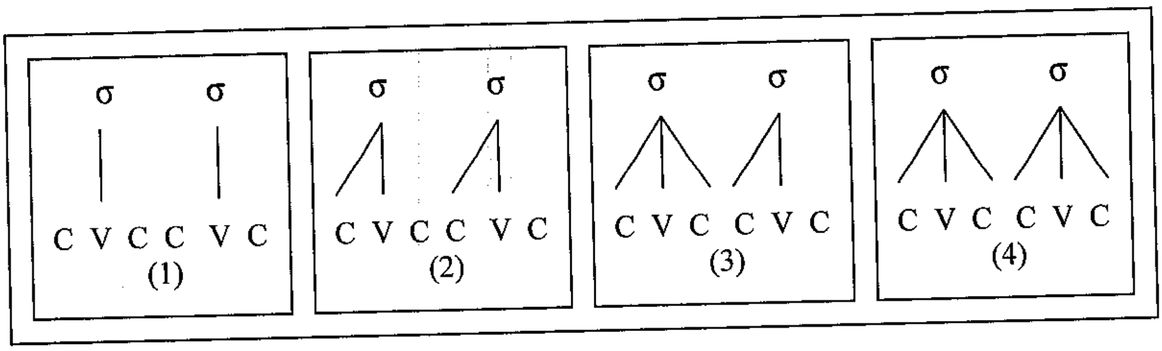


Figure 3.6 Special procedures in MOP process

3.3.4 Speech Production

Speech production is the final stage of the total cycle in producing speech from text. In this stage, the output derived from the previous stage will be mapped to its speech sound accordingly. In order to map the sound to the input, I am using the syllable to sound mapping method.

In this method, several sets of pre-stored audio output are prepared prior to match with the syllable input. These speech sounds were stored in a form of .mp3 files inside a lookup table sound database.

Whenever a polysyllabic or monosyllabic input is given to the system, it will map to the speech sound from the look up sound database accordingly. In order to treat this syllable as an input, each syllable will be named according to the orthographic representation of the syllable. Therefore, it is easier for the matching sound file to be called from the look up table sound database.

To simplify on how the mapping construction works, the syllable to sound mapping is a direct mapping system where correct sound file name will correspond to its correct syllable name.

Hence, each speech sound will follow the same naming convention as the syllable for the system to identify perfect file to be called and produce its sound. Once every syllable is mapped to its corresponding speech sound, each syllable sound will be concatenated together to produce syllable sounding output.

The collection of the speech audio will range from a single vowel sound up to a combination of two consonant per vowel in a syllable following the rules that is identified earlier. Since this purpose of this project is to create a prototype, the collection of probable sound output is limited.

In accordance to that, there might be a condition where, the system is unable to identify correct mapping of the syllable to its corresponding sound. The overall process of mapping the syllable to its corresponding speech output is illustrated as below.

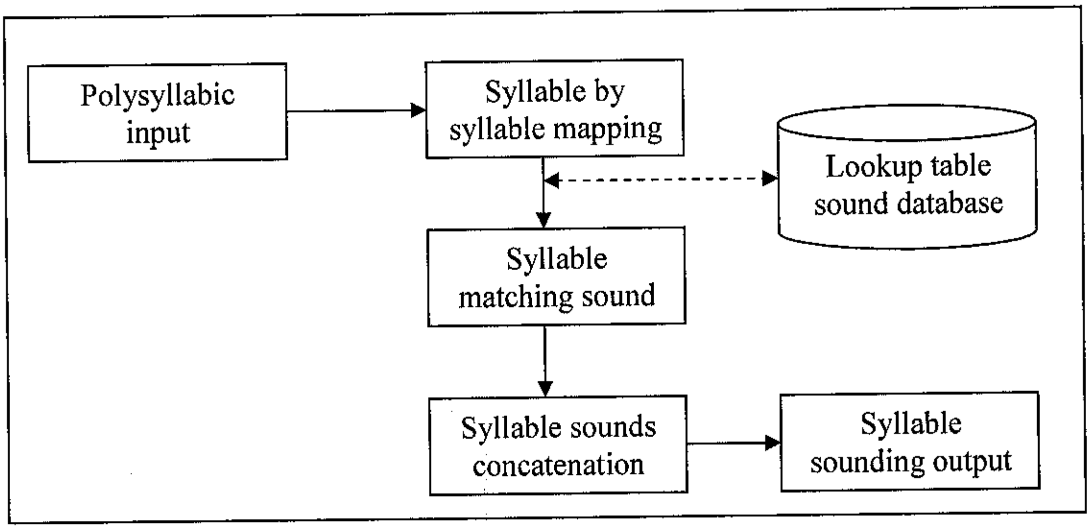


Figure 3.7 Syllable to sound mapping method

3.4 Tools Required

In performing this project several tools and equipments are needed. Following are minimum requirement needed to perform this project and to use the application (hardware requirements).

3.4.1 Hardware Requirements

1. Personal Computer/ Desktop
    - Windows 98, NT, 2000, or XP
    - Pentium 4 processor 1.4 GHz or higher
    - 64 MB RAM memory or higher
    - 1 GB hard disk space or higher

2. Audio Tools
    - Speaker or Headphone
    - Microphone
    - Sound Card

3.4.2 Software Requirements

1. Programming Tool
    - Microsoft Visual Basic 6
2. Audio processing Tool
    - Microsoft Sound Recorder 5.1

CHAPTER 4

RESULT AND DISCUSSION

As the knowledge gathering and system development process is still going on, there are a lot of things that I had found out and accomplished until the current moment. It is pretty much enlightening since there are more questions were answered and more problems seem to have its answer.

4.1 Research Findings

Upon development of this project, the biggest findings were the adoption of Maximum Onset Principle and reflection of unit selection method into text-to-speech processing for Bahasa Melayu.

The maximum onset principle catalyzes the text extraction for the development of this project. It is essential for an input text to be properly normalized and syllabified in order for it to be readable by the system and later assigned to its suitable sound.

Journals from previous researcher and text published by Dewan Bahasa Dan Pustaka helped the extraction of essence of the language structure and its dependencies for it to be implemented with the MOP method. Though MOP is normally used for English, I found that it is acceptable for it to be implemented to Bahasa Melayu due some similarities that both languages are sharing.

It is a major breakthrough once the principle is experimented on Bahasa Melayu's text to test on its reliability. The result shows that MOP is suitable for syllable extraction in the use of Bahasa Melayu. Reason being behind the successfulness of this experiment is Bahasa Melayu share similar syllable structure as English – having to have onset, nucleus and coda in a syllable structure.

Meanwhile, unit selection method had contributed to this project in terms of normalized text to sound production approach. Through research and findings, it has initiated the concept of syllable to sound mapping that is implemented for this project. The idea behind the unit selection that uses pre-stored speech segments created from phonemic representation (phonemes, diphonemes), syllables and acoustic signal information in a database had brought up the syllable to sound mapping concept.

The concept inspired from the unit selection method shares the main principle of having pre-stored speech in a database for closest text processing output mapping. In this case of study, the syllable to sound mapping works as a database of a set of lookup table containing pre-stored syllable sounding speech for it to be mapped to a given post-processed text (i.e. syllables structures).

All procedures outlined for this project might not have worked as it is intended to be without proper understanding on the language and linguistic characteristics. Dewan Bahasa Dan Pustaka provided clear description of the language structures, grammar and its characteristics that leads to the preparation of how should this system work.

The coincidental fact of having similar alphabetic systems and syllable structure helped the development of this project where resources pertaining to English syllable structure, articulations and phonology contributed much in concept comparison and replication.

The project outcome is reflected by the concept and methods mentioned above. The convenience of implying English's TTS concept approaches and theories into Bahasa Melayu very much refined the total development of this project.

4.2 Result Discussion

In meeting with the project requirements, the end product prototype should be able to produce syllable sounding speech from a text insertion. This prototype will process standard Malay text insertion given with several limitations and boundaries.

This system will check the validity of an input before syllabification process is performed on the text. Validity of the text will be based on valid consonant cluster rules where only certain consonant cluster is permissible in a word. From the limitation given to this system, a maximum of 2 consonant clusters will be permitted in a word. Words containing more than 2 consonants will automatically void the operation where user will have to enter a new text.

The system will evaluate the string entered character by character from left to right sequence to examine consonant clusters between each character. Apart from that, the system will also check on non alphabetic symbol such as punctuations, numbers and special characters. These non alphabetic symbols will terminate the process once found in the character sequence (see APPENDIX C for examples).

Word	Maximum Consonant Clusters	Validity
Bacang	**2 – NG (end of word)**	Valid
Nganga	**2 – NG (beginning and middle of word)**	Valid
Murni	**2 – RN (midle of word)**	Valid
Kenchana	**3 – NCH (middle of word**	Invalid – beyond maximum value
Tmpatanng	**3 – TMP, NNG (beginning and middle of word)**	Invalid – beyond maximum value

Table 4.0 Valid consonant clusters

Word	Non-alphabetic symbol	Reason
Baw22ang	Numbers – **22**	Invalid – Numbers used in word
Ka.cip	Punctuation – **'.' or full stop sign**	Invalid – Punctuation used in word
Ya&tidak	Special character – **'&' or and sign**	Invalid – Special character used in word

Table 4.1 Invalid Non-alphabetic symbols

This validity check however will not determine the correct spelling of a standard Malay word or determine any meaning to the word entered. The reason behind this is to minimize the risk of having erroneous result so the text will be properly sounded. Valid inputs will be eligible for syllabification process into its syllable structure by using syllable boundaries to separate one syllable to another.

The text inserted will be syllabified using 3 rules extracted from the Dewan Bahasa Dan Pustaka syllabification rules which are the diphthong rules, permissible onset sequence rules and permissible consonant clusters rules. Each rule will help the syllabification process for determining syllable boundary for given word. In concurrent with these rules is the adoption of MOP as the algorithm of structuring the syllables.

Words that had been syllabified using MOP will be mapped to its closest speech sound from the lookup table located in the database. The lookup table consist limited number of speech sound combination due to the fact that the purpose is to test whether it is reliable and its workability in a small scale application.

Therefore, there might be a combination of a syllable that will not have appropriate speech sound mapped to it and this might cause inconsistencies in speech results. Though, I am positive that as long as the basic requirement of this system is made, the chances of facing this problem are very low.

The system might also face unwanted outcome or speech production failure if the words entered were loanwords and acronyms. This is because the system could not differentiate the loanwords from the basic Bahasa Melayu words. Thus, it is inevitably that the system would not work to its fullest extend if such condition apply. However, it is reminded in the user manual that loanwords apart from other limitation were outline as the system limitation.

In accordance to that, there might be some loanwords that might slip through the system due to its syllable pattern. Word such as "boikot" (boycott), "focus" (focus) and "fungsi" (function) are some of loanwords that will get through the system. Some might be able to be pronounce if they follow the syllable pattern or sequence outline earlier and some might not because of incompatibility to the sequences predefine before (see APPENDIX C for more examples).

Finally, there is an issue of different phonemic representation of character 'e' although it was orthographically written the same. The character 'e' can be pronounced in two ways denoted by 'e taling' and 'e pepet'. 'e taling' will carry the sound of vowel /e/ as the pronunciation of the word "enak" (delicious) while 'e pepet' will carry the sound of schwa /δ/ (Yousif A. El-Imam, Zuraida Mohammed Don, 2000) as the pronunciation of the word "emas" (gold).

Since majority of Bahasa Melayu words that have "e" in their orthography carry the schwa pronunciation (Yousif A. El-Imam, Zuraida Mohammed Don, 2000), I had decided to use an schwa pronunciation to all words that contain "e" to resolve ambiguity between 'e taling' and 'e pepet'. Thus, there will be a difference in speech sound production to words that follow the /e/ pronunciation.

CHAPTER 5

CONCLUSION

5.1 Conclusion

This project contributes methods to automating written text through syllabification process for Bahasa Melayu and produce syllable sounding speech from it. The examined literature review discovered that no attempt has been made in creating a TTS system for Bahasa Melayu and the usage of Maximum Onset Principle as part of the syllabification system. This project illustrated the feasibility and accuracy of performing syllabification using the MOP method in producing correct syllable structure for it to be use as an input to speech prosody.

Also illustrated in this project is the reflection of unit selection method in creating syllable to speech mapping concept in mapping closest speech sound to a syllable extracted from the syllable structure. This project also discussed successfulness in the method used within the limitation given for the system to prove that TTS conversion is possible to be done for Bahasa Melayu using existent concepts and principles.

Finally, from the research and development done for this project, I can fairly claimed that I have reached the objectives that has been stated earlier and hope that with the development and study performed will later become a stepping stone for an advance level of text-to-speech conversion that uses Bahasa Melayu and perhaps able to promote the language itself inadvertently as a new language for knowledge seekers.

5.2 Future Development Recommendation

On a personal note, I do not see this project as fully completed due to several limitation made to the system. Perhaps for future development, this system can be expanded in treating loanwords into the system. Apart from that, it would be interesting if the ambiguity between the /e/ and /ə/ pronunciation to be solved in future development.

Furthermore, the system will be best to its standard if it can produce natural sound speech by having smooth blending between syllable sounding and cater for sonority of the speech sound with the assistance of punctuations and syllable structure model.

REFERENCES

1. A. S. Hornby. 1995, *Oxford Advance Learner's Dictionary of Current English.* Oxford University Press. Wwalton Street. Oxford.

2. Caroline Floccia, Jeremy Goslin*, Naanaa Bouketir, Joel Bradmetz. 1999, *Consonant Cluster Syllabification in Preliterate Children.* Laboratory of Psychology, Universite de Franche-Comte, France. (*) Laboratory of Experimental Psycholinguistics, Universite de Geneve, Switzerland and Department of Computer Science, University of Sheffield, UK.

3. Hajah Noresah bt. Baharom. 1994. *Kamus Dewan.* Edisi Ketiga. Dewan Bahasa Dan Pustaka, Kementerian Pendidikan Malaysia, Kuala Lumpur.

4. Joran Meron. 2001, *Prosodic Unit Selection Using an Imitation Speech Database.* Panasonic Technologies, Inc., Panasonic Speech Technology Lab., Santa Barbara, CA.

5. Ksenia Shalonova, Roger Tucker. Jan 2003, *South Asian Languages in Multilingual TTS-Related Database.* Mobile and Media System Laboratory, HP Lab, Bristol, U.K.

6. Li Zhang, William H Edmondson. 2002, *Speech Recognition Using Syllable Patterns.* School of Computer Science, University of Birmingham, Birmingham, UK.

7. *Macmillan English Dictionary for Advance Learner.* 2002. Macmillan Publishers Limited, Oxford, UK.

8. Nik Safiah Karim. 1995, *Malay Grammar for Academics and Professionals.* Dewan Bahasa Dan Pustaka, Kementerian Pendidikan Malaysia, Kuala Lumpur.

9. Nik Safiah Karim, Farid M. Onn, Hashim Hj. Musa, Abdul Hamid Mahmod. 1996, *Tatabahasa Dewan.* Edisi Tebaru. Dewan Bahasa Dan Pustaka, Kementerian Pendidikan Malaysia, Kuala Lumpur.

10. Paul C. Bagshaw, Briony J. Williams. 1992, *Criteria for labeling prosodic aspects of English Speech.* Centre for Speech Technology Research, University of Edinburgh, Scotland, UK.

11. Romain Prudon, Christophe d'Alessandro. 2002, *A selection/concatenation text-to-speech synthesis system: database development, system design, comparative evaluation.* LIMSI-CNRS. BP133 – Universite Paris XI, F91403 Orsay, France.

12. Roxanne Ruzic and Kathy O'Connell, *Text-to-Speech - National Center on Accessing the General Curriculum.* Center for Applied Special Technology, CAST.

13. T. Lambert, A.P. Breen, B. Eggleton, S.J. Cox, B.P. Milner. 2003, *Unit Selection in Concatenative TTS Synthesis Systems Based on Mel Filter Bank Amplitude and Phonetic Context.* University of East Anglia, Nuance Communication, Norwich.

14. Thierry Dutoit. 1996, *High-Quality Text-to-Speech Synthesis: an Overview, A Short Introduction to Text-to-Speech Synthesis.* Journal of Electrical & Electronics Engineering, Australia: Special Issue on Speech Recognition and Synthesis vol. 17 n°1.

15. Thomas Russi. 1990, A *syntactic and morphological analyzer for a text-to-speech system.* International Conference on Computational Linguistics archive. Proceedings of the 13th conference on Computational linguistics – Volume 3, Swiss Federal Institute of Technology (ETH), Zürich.

16. William H Edmondson, Li Zhang. 2002, *The Use of Syllable Structure for Speech Recognition*. School of Computer Science, University of Birmingham, Birmingham, UK.

17. Yoshifumi Ooyama, Masahiro Miyazaki, Satoru Ikehara. 1987, *Natural Language Processing In A Japanese Text-to-Speech System*. NNT Communication and Information Processing Laboratories, 1-2356 Take, Yokosuka, Kanagawa-ken, Japan.

18. Yousif A. El-Imam, Zuraida Mohammed Don. Feb 2000, *Text-to-Speech Conversion of Standard Malay*. International Journal of Speech Technology.

19. *Bahagian Satu. Pedoman Umum Ejaan Rumi Bahasa Melayu*.2004, Dewan Bahasa Dan Pustaka.
http://www.karyanet.com.my/knet/dbp_jawi/pengenalan/1-01_latarbelakang.htm

20. *Dictionary.com*
http://dictionary.com

21. *How Text-to-Speech Works*
http://project.uet.itgo.com/textto1.htm

22. Hyperdictionary.
http://www.hyperdictionary.com.

23. Nationmaster.com
http://www.nationmaster.com/encyclopedia/Bahasa-Melayu

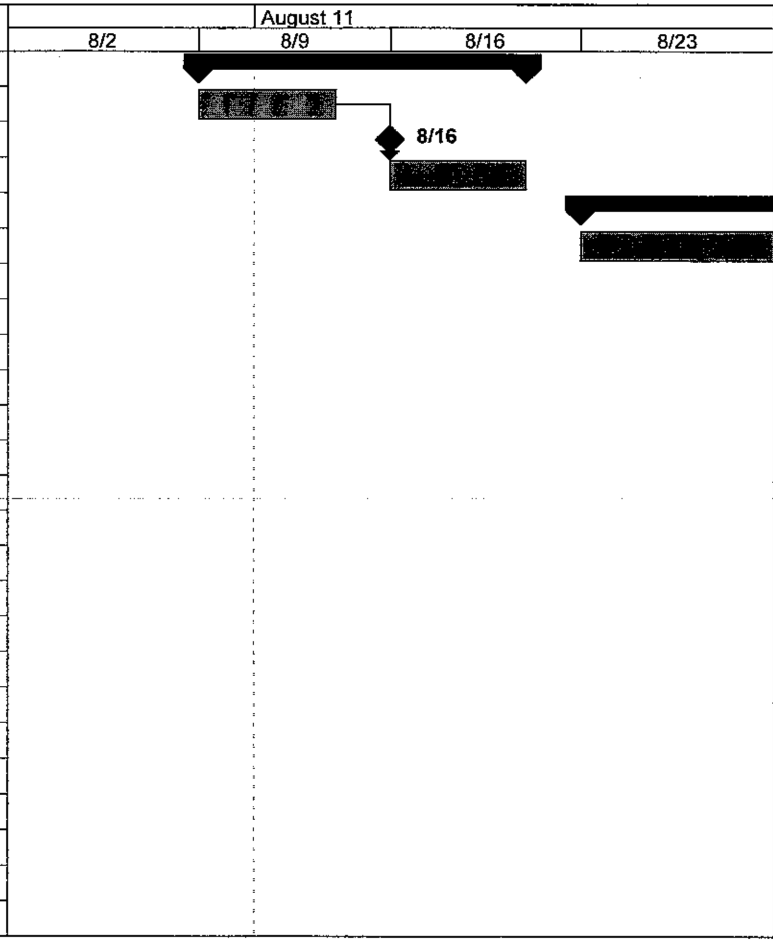
24. What are syllables?
http://www.ic.arizona.edu/~lsp/Phonology/Syllables/Phonology3.html#

25. WIKIPEDIA, The Free Encyclopedia. *Speech synthesis.*
http://en.wikipedia.org/wiki/Text_to_speech

APPENDICES

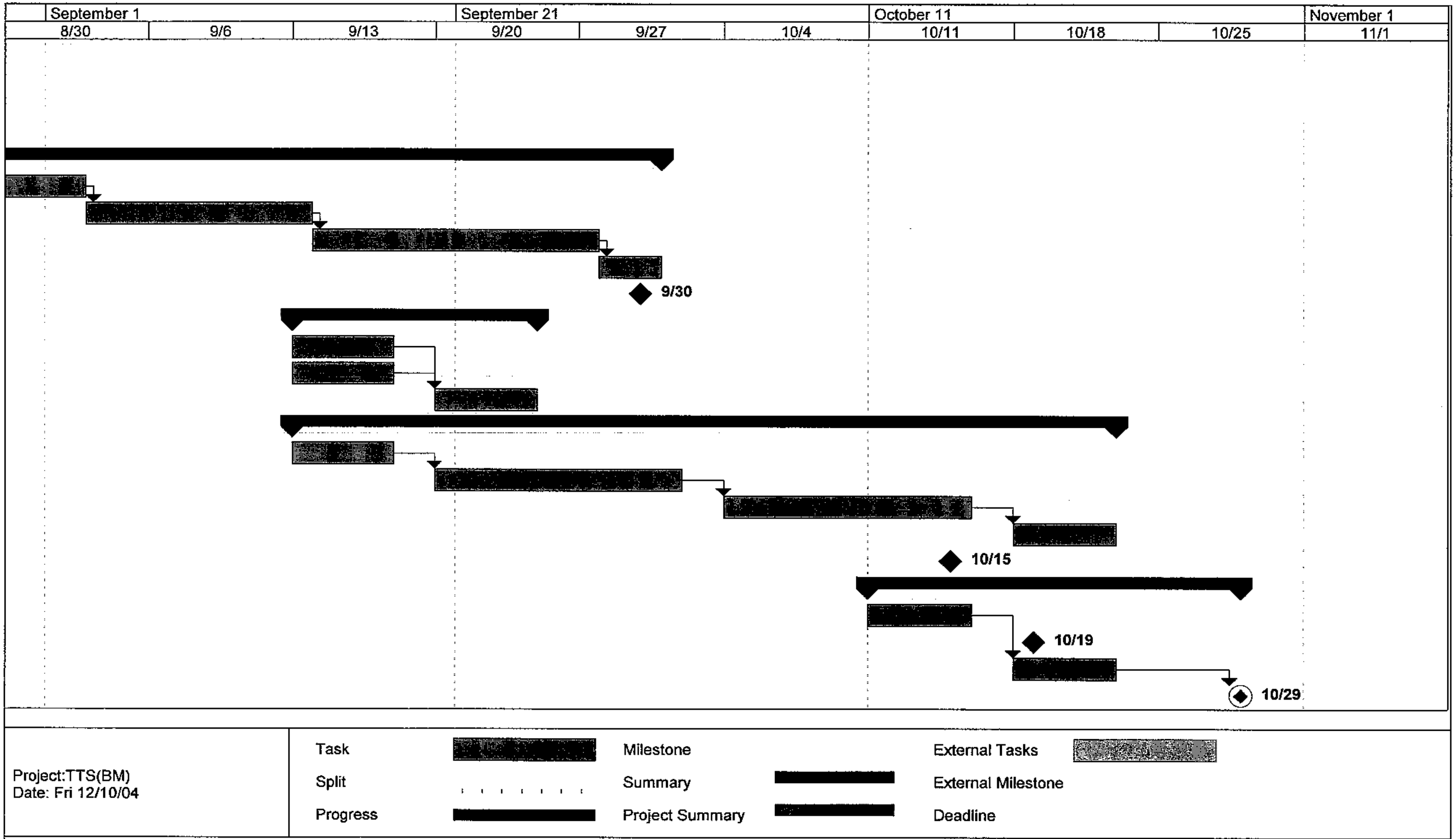
APPENDIX A: PROJECT TIMELINE/ GANTT CHART

ID		Task Name	Duration	Start	Finish
1		**1.0 System Initiation and Feasibility Stu**	**10 days**	**Mon 8/9/04**	**Fri 8/20/04**
2		1.1 Proposal Submission	1 wk	Mon 8/9/04	Fri 8/13/04
3		1.2 Preliminary Report Submission	0 days	Mon 8/16/04	Mon 8/16/04
4		1.3 Project Sceduling	1 wk	Mon 8/16/04	Fri 8/20/04
5		**2.0 Project Planning and Funtional Anal**	**29 days**	**Mon 8/23/04**	**Thu 9/30/04**
6		2.1 Data Gathering	1.8 wks	Mon 8/23/04	Thu 9/2/04
7		2.2 Data Analysis	1.4 wks	Fri 9/3/04	Mon 9/13/04
8		2.3 Method Generation	10 days	Tue 9/14/04	Mon 9/27/04
9		2.4 Requirement Specification	0.6 wks	Tue 9/28/04	Thu 9/30/04
10		2.5 Progress Report Submission	0 days	Thu 9/30/04	Thu 9/30/04
11		**3.0 System Design**	**10 days**	**Mon 9/13/04**	**Fri 9/24/04**
12		3.1 Interface Design	1 wk	Mon 9/13/04	Fri 9/17/04
13		3.2 System Design Modelling	1 wk	Mon 9/13/04	Fri 9/17/04
14		3.3 Data Structure Design	1 wk	Mon 9/20/04	Fri 9/24/04
15		**4.0 Programming**	**30 days**	**Mon 9/13/04**	**Fri 10/22/04**
16		4.1 Interface Development	1 wk	Mon 9/13/04	Fri 9/17/04
17		4.2 Algorithm Generation	2 wks	Mon 9/20/04	Fri 10/1/04
18		4.3 System Programming	2 wks	Mon 10/4/04	Fri 10/15/04
19		4.4 System Intergration	1 wk	Mon 10/18/04	Fri 10/22/04
20		4.5 SV Final Draft Submission	0 days	Fri 10/15/04	Fri 10/15/04
21		**5.0 Implementation Phase**	**14 days**	**Mon 10/11/04**	**Fri 10/29/04**
22		5.1 System Implementation	1 wk	Mon 10/11/04	Fri 10/15/04
23		5.2 Product Demonstartion	0 days	Tue 10/19/04	Tue 10/19/04
24		5.3 System Delivery	1 wk	Mon 10/18/04	Fri 10/22/04
25		5.4 Final Delivery	0 days	Fri 10/29/04	Fri 10/29/04



Project:TTS(BM)
Date: Fri 12/10/04

Task | Milestone | External Tasks
Split | Summary | External Milestone
Progress | Project Summary | Deadline


September 1
September 21
October 11
November 1
8/30
9/6
9/13
9/20
9/27
10/4
10/11
10/18
10/25
11/1
9/30
10/15
10/19
10/29
Project:TTS(BM)
Date: Fri 12/10/04
Task
Split
Progress
Milestone
Summary
Project Summary
External Tasks
External Milestone
Deadline

APPENDIX B: SYSYEM REQUIREMENT AND USER GUIDE

APPENDIX B

System Requirement and User Guide

The minimum requirements to operate the program are listed below:

Hardware requirements:

1. Personal Computer/ Desktop
   - Windows 98, NT, 2000, or XP
   - Pentium 4 processor 1.4 GHz or higher
   - 64 MB RAM memory or higher
   - 1 GB hard disk space or higher
2. Audio Tools
   - Speaker or Headphone
   - Sound Card

System Limitation:

1. This program will only read by syllable.
2. The maximum word to be entered is one word.
3. The input to the system at present is only a piece of standard Bahasa Melayu text and will not process a text in batch or a text file.
4. The input can consist of monosyllabic words but will not be able to process words with syllabic consonants.
5. This system will not cater loanwords, punctuations, abbreviations, acronym and numbers.
6. The input will need to have a maximum of 2 consonant clusters in a given word regardless of it preposition to ensure smooth text to sound processing.
7. Since this is a prototype, it will only produce up to certain capacity of result.

(See list of loanwords, punctuations, abbreviations, acronym and numbers in APPENDIX)

How to Use This Program

1. From the CD provided, copy the sound folder into your **C:** \ drive.
2. Run the TTS BM execution file from the CD.
3. The program is successfully executed and ready for use.

APPENDIX C: EXAMPLE OF LOANWORDS, PUNCTUATIONS, ABBREVIATION, ACRONYM AND NUMBERS

APPENDIX C

Examples of Loanwords, Punctuations, Abbreviations, Acronym and Numbers

Source: Dewan Bahasa Dan Pustaka (various publications and write up)

Nationmaster.com

http://www.nationmaster.com/encyclopedia/Bahasa-Melayu

Bahasa Melayu Loanwords

English

Original Word	Converted Word
Race	"ras"
Abstract	"abstrak"
Classic	"klasik"
Effective	"efektif"
Cylinder	"silinder"
Machine	"mesin"
Dhobi	"dobi"
System	"sistem"
Jeep	"jip"
Stereo	"stereo"
Factor	"faktor"
Frequency	"frekuensi"
Exclusive	"eksklusif"
Dialogue	"dialog"
Quota	"kuota"
Extra	"ekstra"
University	"universiti"
Modernism	"modenisme"

APPENDIX C (cont'...)

Nusantara Region

Original Word	Converted Word
Gembleng	"gembleng"
Jempol	"jempol"
Pamer	"pamer"
Kaget	"kaget"
Lelucon	"lelucon"
Perabot	"perabot"
Awet	"awet"
Lapor	"lapor"
Bunglon	"bunglon"
Larik	"larik"
Payon	"payon"

Arabic

Original Word	Converted Word
Dharurat	"darurat"
Hadhir	"hadir"
Dharab	"darab"
Ghairah	"ghairah"
Maghrib	"maghrib"
Mubaligh	"mubaligh"
Khabar	"khabar"
Akhir	"akhir"
Khusus	"Khusus"
Sharia	"Syariah"
Duniyya	"dunia"

APPENDIX C (cont'...)

Sanskrit

Original Word	**Converted Word**
Abda	"sabda"
Astra	"sastera"
Bahasa	"bahasa"
Roti	"roti"

Chinese (Hokkien)

Original Word	**Converted Word**
Beehoon	"bihun"
Mee	"mee/mi"
Tauhu	"tauhu"
Teh	"teh"
Teh ko	"teko"

Portuguese

Original Word	**Converted Word**
Bombeiros	"bomba"
Igreja	"gereja"
Queijo	"keju"
Lim o	"limau"
Manteiga	"mentega"

APPENDIX C (cont'...)

Punctuations

Name	Symbol
Full stop	(.)
Comma	(,)
Semicolon	(;)
Colon	(:)
Dash/hyphen	(-)
Ellipse	(...)
Question mark	(?)
Exclamation mark	(!)
Brackets	(( ))
Inverted comma	(“ ”)
Quote	(‘ ’)
Slash	(/)
Apostrophe	(`)

APPENDIX C (cont'...)

Abbreviation

Abbreviation	Meaning
sbg.	sebagai
dll.	dan lain-lain
prof.	profesor
pn.	puan
kol.	kolonel
sdr.	saudara
dr.	doktor
en.	encik
tn.	tuan
Y.A.B	yang amat berhormat
Y.M	yang mulia

Acronym

Acronym	Meaning
JKR	Jabatan Kerja Rakyat
KDN	Kementerian Dalam Negeri
Pernas	Perbadanan Nasional
PETRONAS	Petroliam Nasional
TLDM	Tentera Laut Diraja Malaysia
UMNO	United Malayan National Organization
WHO	World Health Organization

APPENDIX C (cont'...)

Numbers

Modern numbering: **0, 1, 2, 3, 4, 5, 6, 7, 8, 9**

Roman numbering: **I, II, III, IV, V, VII, VIII, IX, X, L (50), C (100), D (500), M(1000), V(5000), M (1000000)**

APPENDIX D: INTERFACE DESIGN – STORY BOARD AND DESCRIPTIONS

APPENDIX D

Interface design process and iterations.

Stage One – Generalizing Concept and Ideas

Preliminary interface design (sketches)

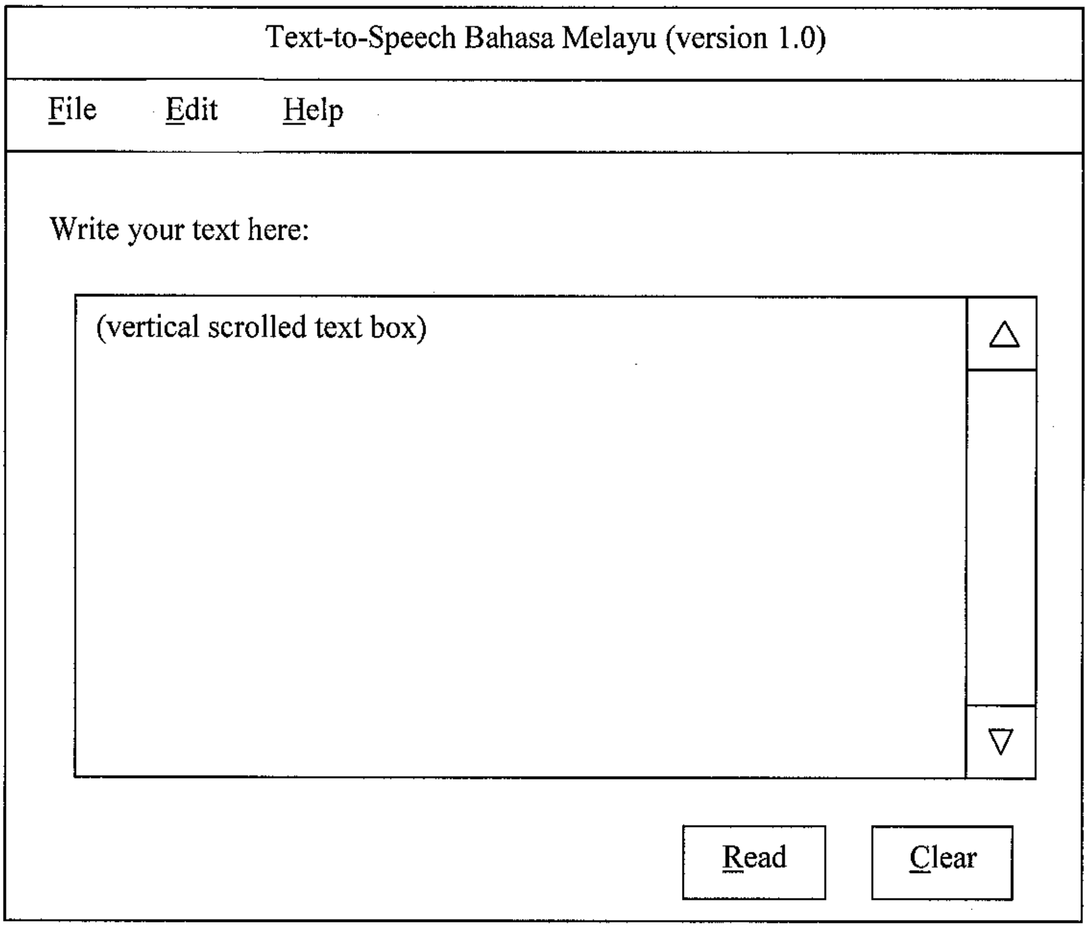


Illustration 1 First draft of the interface design (main form)

APPENDIX D (cont' ...)

Description:

1. Title – TTS Bahasa Melayu (version 1.0)
2. Menu1 – File
    - Submenu1 – New
    - Submenu2 – Open
    - Submenu3 – Close
    - Submenu4 – Save
    - Submenu5 – Save As
    - Submenu6 - Exit
3. Menu2 – Edit
    - Submenu1 – Copy
    - Submenu2 – Paste
4. Label – Text: Write your text here:
5. Vertical scrolled textbox – Text (none)
6. Button1 – Read
7. Button2 – Clear

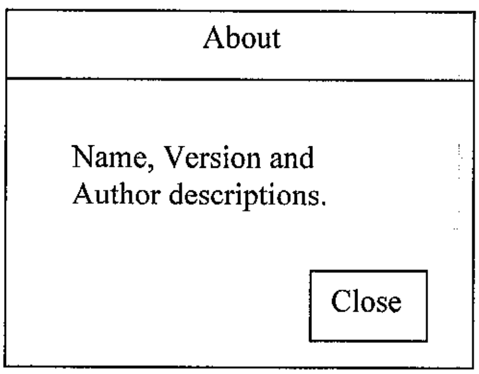


Illustration 2 First draft of the interface design (sub form)

APPENDIX D (cont' ...)

Stage Two – Refining and Design Evaluation

Final interface design (sketches)

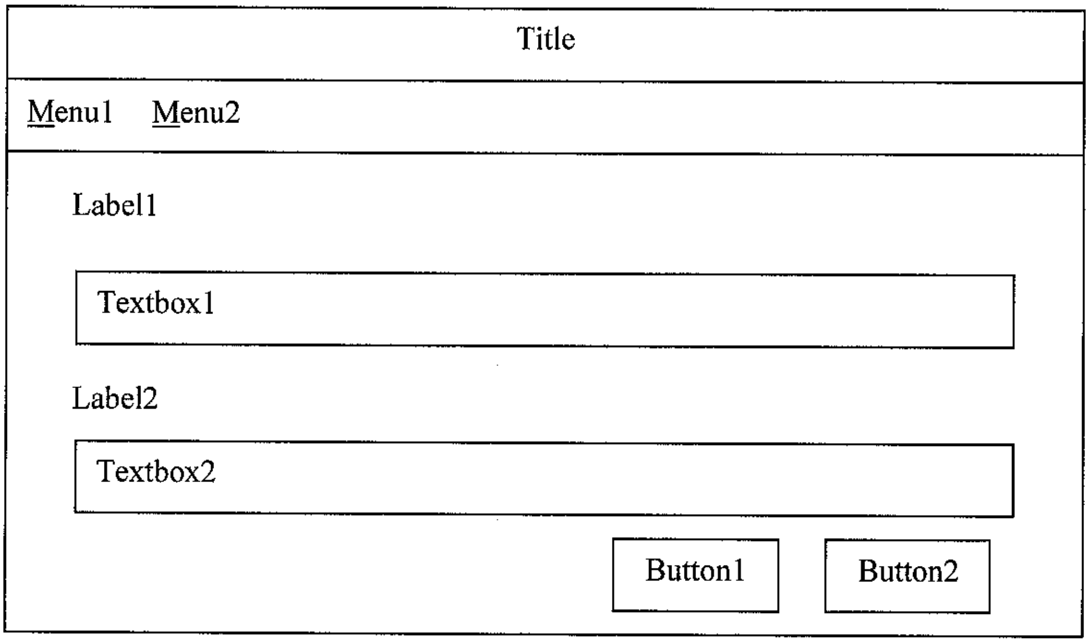


Illustration 3 Final draft of the interface design (main form)

Description:

8. Title – TTS Bahasa Melayu (version 1.0)
9. Menu1 – File
    - Submenu1 – Exit
10. Menu2 – Help
    - Submenu1 – About
11. Label – Text: Please enter your text here:
12. Label – Syllabified Input
13. Textbox1 – Text (default): Cubaan
14. Textbox2 – Text output
15. Button1 – Speak
16. Button2 – Clear

APPENDIX D (cont' ...)

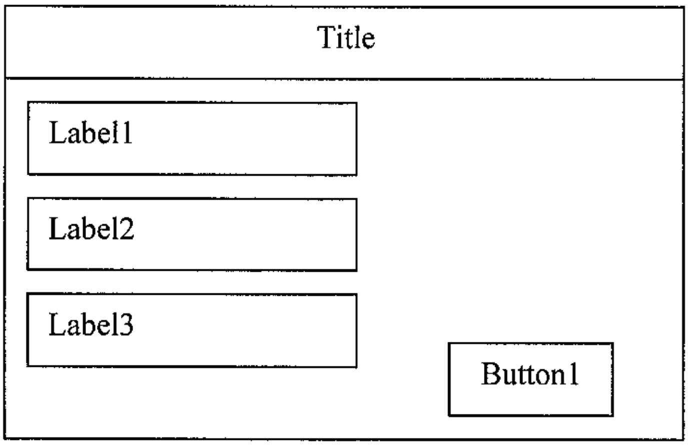


Illustration 4 Final draft of the interface design (sub form)

Description:

1. Title – About: TTS Bahasa Melayu
2. Label1 – Name of application
3. Label2 – Further description of application and version
4. Label3 – Author's name
5. Button1 – Close

APPENDIX E: INTERFACE DESIGN - SNAPSHOT

APPENDIX E

Interface Design Snapshot

<u>Stage Three – Real Design</u>

Sketches implementation (image snapshots)

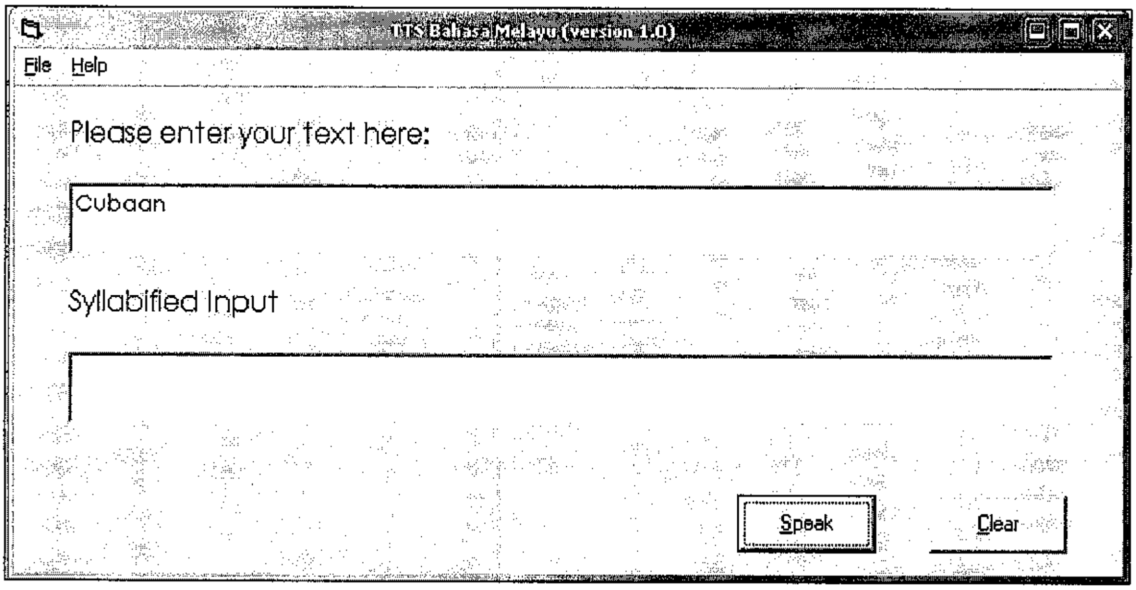


Figure 1 Interface design (main form)

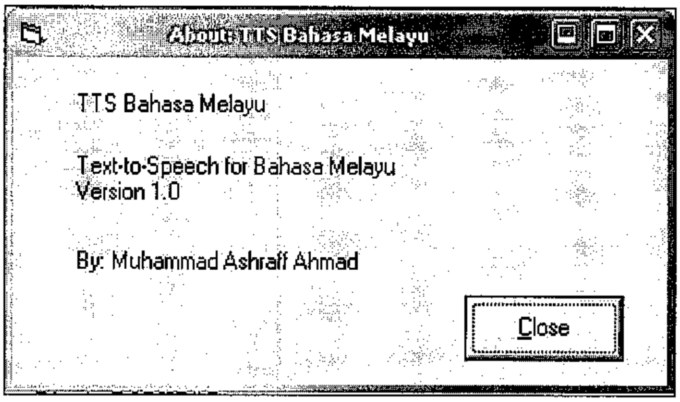


Figure 2 Interface design (sub form)

Sample of Results

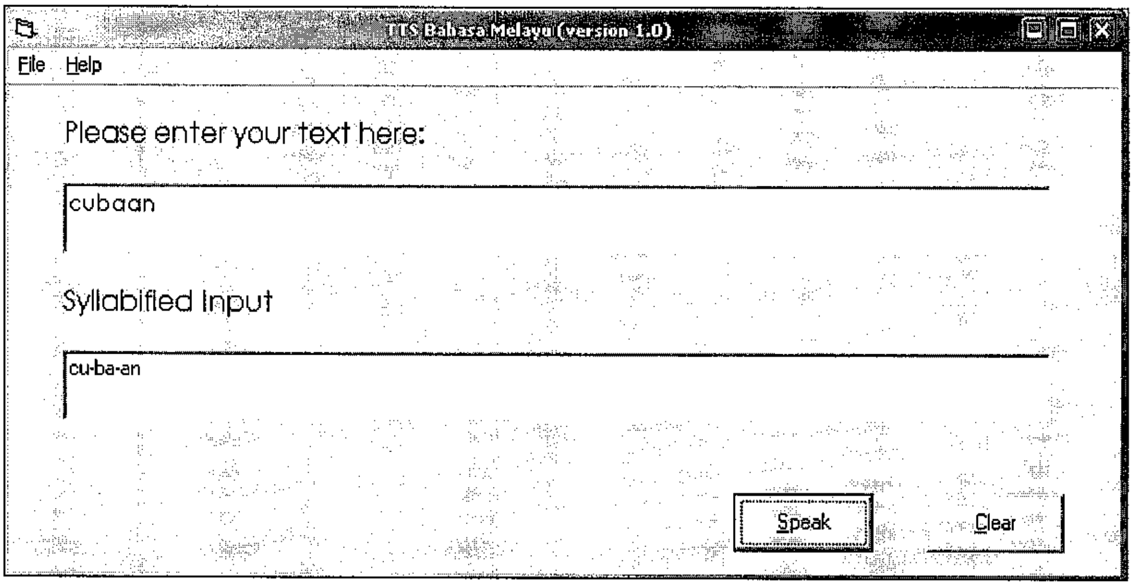


Figure 3 Syllabification for word "cubaan" (testing)

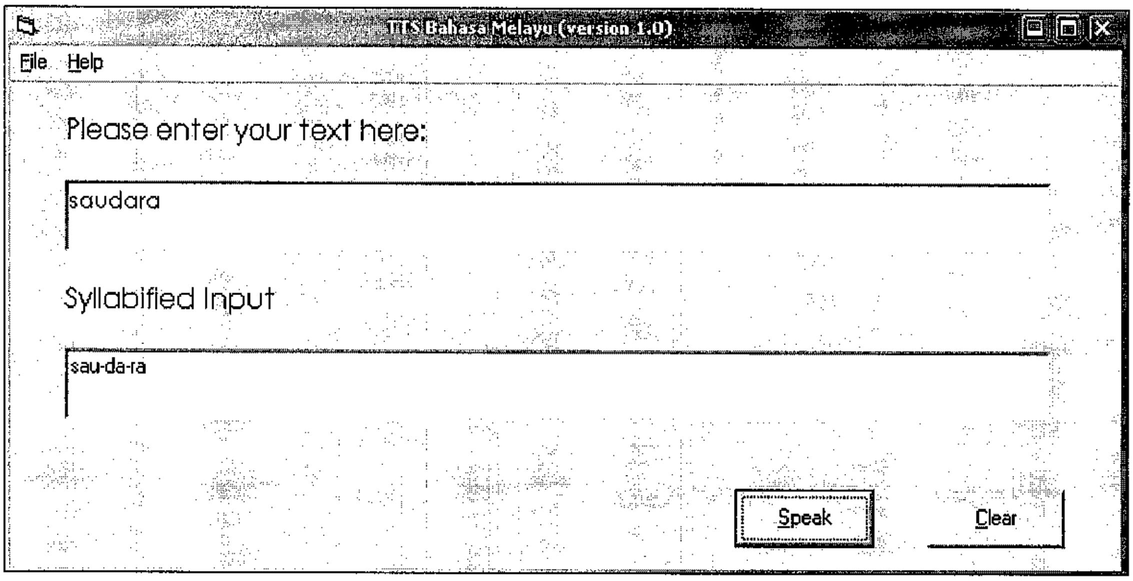


Figure 4 Syllabification for word "saudara" (relative)

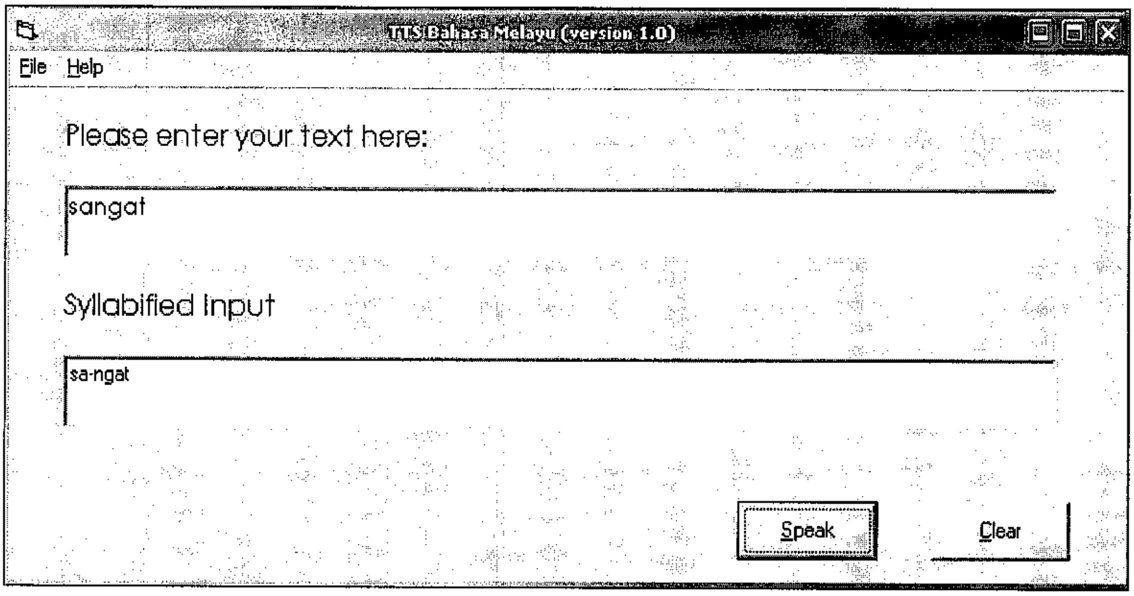


Figure 5 Syllabification for word "sangat" (very)

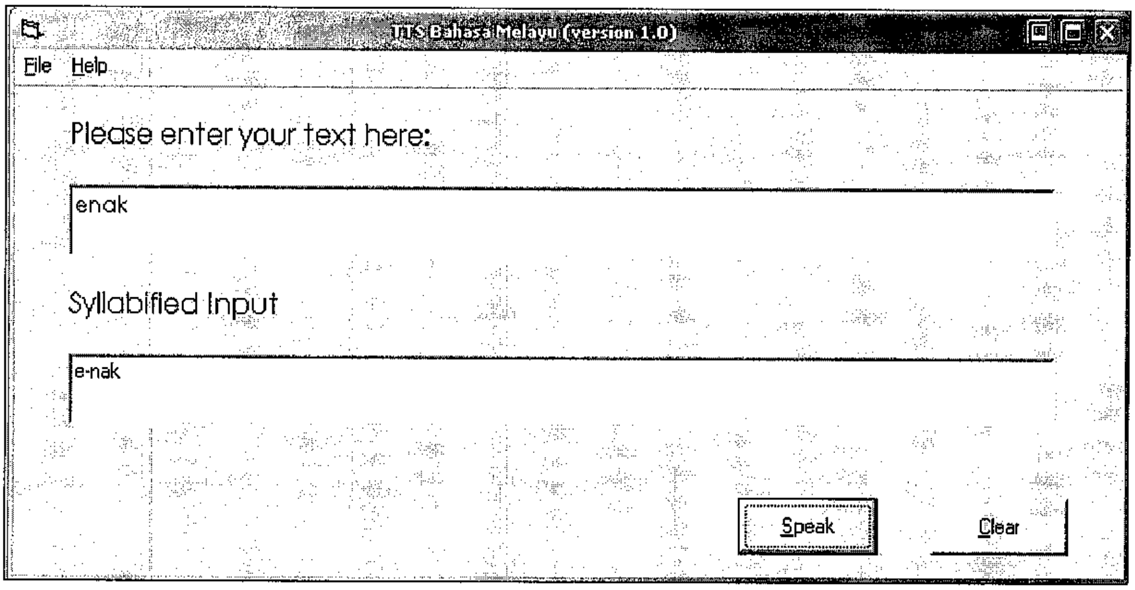


Figure 6 Syllabification for word "enak" (delicious)