# BOOTSTRAPPING KELANTAN AND SARAWAK MALAY DIALECT MODELS ON TEXT AND PHONETIC ANALYSES IN TEXT-TO-SPEECH SYSTEM

**JASMINA KHAW YEN MIN**

**UNIVERSITI SAINS MALAYSIA**

**2017**

# BOOTSTRAPPING KELANTAN AND SARAWAK MALAY DIALECT MODELS ON TEXT AND PHONETIC ANALYSES IN TEXT-TO-SPEECH SYSTEM

**by**

**JASMINA KHAW YEN MIN**

**Thesis submitted in fulfillment of the requirements**

**for the Degree of**

**Doctor of Philosophy**

**July 2017**

## DECLARATION

Name: **JASMINA KHAW YEN MIN**

Matric No: **P-COD0010/12(R)**

School: **SCHOOL OF COMPUTER SCIENCES**

Thesis Title: **BOOTSTRAPPING KELANTAN AND SARAWAK MALAY DIALECT MODELS ON TEXT AND PHONETIC ANALYSES IN TEXT-TO-SPEECH SYSTEM**

I hereby declare that this thesis I have submitted to **SCHOOL OF COMPUTER SCIENCES** on **JULY 2017** is my own work. I have stated all references used for the completion of my thesis.

I agree to prepare electronic copies of the said thesis to the external examiner or internal examiner for the determination of amount of words used or to check on plagiarism should a request be made.

I make this declaration with the believe that what is stated in this declaration is true and the thesis as forwarded is free from plagiarism as provided under Rule 6 of the Universities and University Colleges (Amendment) Act 2008, University Science Malaysia Rules (Student Discipline) 1999.

I conscientiously believe and agree that the University can take disciplinary actions against me under Rule 48 of the Act should my thesis be found to be the work or ideas of other persons.

Students Signature: ………………………………… Date: ………………………

Acknowledgement of receipt by: …………………… Date: ………………………

## ACKNOWLEDGEMENT

Firstly, I would like to express my sincere gratitude to my supervisor Dr. Tan Tien Ping for the continuous support of my PhD study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better supervisor for my PhD study.

Besides my supervisor, I would like to thank my co-supervisors of Associate Professor Dr. Tan Enya Kong and Associate Professor Dr. Ranaivo-Malancon Bali, for their insightful comments and encouragement, but also for the hard question which incented me to widen my research from various perspectives.

I would like to take the opportunities to thank again all the volunteers that have participated in the data collection process and perception tests. Last but not least, I would like to thank my family and friends that are always by my side, motivate me and help me, to make my work possible.

## TABLE OF CONTENTS

**CHAPTER 2-RESEARCH BACKGROUND**



**CHAPTER 3-TEXT ANALSIS AND PHONETIC ANALYSIS IN TEXT-TO-SPEECH SYSTEM**

**CHAPTER 5-EXPERIMENTS AND RESULTS**



**CHAPTER 6-DISCUSSION**



**CHAPTER 7-CONCLUSION AND FUTURE WORKS**

## LIST OF TABLES

## LIST OF FIGURES

**Page**

## LIST OF SYMBOLS AND ABBREVIATIONS

| | |
|---|---|
| G2G | Grapheme to Grapheme |
| G2P | Grapheme to Phoneme |
| HMM | Hidden Markov Model |
| IPA | International Phonetic Alphabet |
| TTS | Text-to-speech |
| BLEU | Bilingual Evaluation Understudy |
| VOT | Voice-Onset Time |
| MOS | Mean Opinion Score |

**PEMODELAN IKAT-BUT DIALEK MELAYU PADA ANALISIS TEKS DAN ANALISIS FONETIK DALAM SISTEM TEKS KEPADA PERTUTURAN**

## ABSTRAK

Teknologi teks kepada pertuturan (TKP) telah matang dan ia telah dilengkapi dan dibenam dalam banyak alat, misalnya dalam telefon bimbit, robotik dan sistem telefoni. Untuk membina sistem TKP, sumber-sumber yang diperlukan ialah korpus ucapan, kamus sebutan dan teks. Walau bagaimanapun, pembinaan sistem TKP dialek Melayu bawah-sumber adalah mencabar. Masalah sumber bahasa yang terhad dalam dialek Melayu termasuk tiada kamus sebutan dialek Melayu, seakan-akan set fonem dialek Melayu yang tidak diketahui, ketiadaan ortografi piawai dalam dialek Melayu, dan teks dialek Melayu bertulis yang terhad. Sistem TKP melibatkan analisis teks, analisis fonetik, analisis prosodik dan sintesis pertuturan. Kajian kami memberi tumpuan kepada analisis teks dan analisis fonetik bagi dialek Melayu bawah-sumber. Dalam tesis ini, satu kerangka untuk TKP bagi dialek Melayu bawah-sumber telah dicadangkan. Selain itu, kami juga mencadangkan pemodelan dialek Melayu secara ikat-but dengan menggunakan sumber Melayu dan sumber berbilang bahasa untuk pembangunan sistem TKP dialek Melayu. Dalam kerangka cadangan kami, pendekatan separa menyelia untuk menterjemahkan korpus bahasa Melayu kepada korpus dialek Melayu telah dicadangkan untuk menyelesaikan masalah teks dialek Melayu bertulis yang terhad. Dalam masalah ini, kami mencadangkan satu pendekatan penjajaran kata dan frasa untuk mendapatkan kosa kata dan peraturan terjemahan dialek Melayu untuk terjemahan mesin. Keputusan menunjukkan bahawa nilai kepersisan dan perolehan kembali lebih daripada 95%.

Seterusnya, algoritma normalisasi yang baru telah dicadangkan untuk menyeragamkan ortografi teks dialek Melayu. Untuk membina TKP system, teks yang seragam diperlukan. Keputusan menunjukkan bahawa perbendaharaan kata dialek telah berjaya diseragamkan berdasarkan ortografi bahasa Melayu dengan ketepatan yang tinggi. Untuk mencari set fonem dialek Melayu, pendekatan pengenalan fonem secara automatik dengan menggunakan sumber berbilang bahasa telah dicadangkan. Keputusan menunjukkan bahawa empat puluh tujuh fonem dialek Kelantan dan tiga puluh empat fonem dialek Sarawak telah didapati. Dengan maklumat fonem dan pemetaan grafem kepada fonem, pendekatan grafem kepada fonem (GKP) dialek Melayu berasaskan peraturan kemudiannya dicadangkan untuk menjana kamus sebutan dialek Melayu. Keputusan eksperimen menunjukkan bahawa ketepatan fonem dan ketepatan perkataan mencapai peratusan melebihi 90%. Kami menilai kerangka sistem TKP dialek Melayu yang dicadangkan melalui ujian persepsi. Keputusan menunjukkan bahawa kualiti pertuturan dialek Melayu yang disintesis dengan menggunakan sistem sintesis pertuturan berasaskan Model Markov tersembunyi melalui kerangka yang dicadangkan menunjukkan bahawa dari skala 1 (buruk) hingga 5 (cemerlang), purata penilaian diberi adalah skala 4, iaitu "baik.

**BOOTSTRAPPING MALAY DIALECT MODELS ON TEXT ANALYSIS AND PHONETIC ANALYSIS IN TEXT-TO-SPEECH SYSTEM**

## ABSTRACT

Text-to-speech (TTS) technologies have matured and they have been equipped and embedded in many tools, for instance in mobile phones, robotics and telephony system. For building a TTS system, resources that required are speech corpus, pronunciation dictionary and text. Nevertheless, building an under-resourced Malay dialect TTS system is challenging. The problems of limited language resources in Malay dialects include inexistence of Malay dialects pronunciation dictionaries, quasi-unknown of the sets of Malay dialects phonemes, no standard orthography in Malay dialects and limited written text of Malay dialects. TTS system involves text analysis, phonetic analysis, prosodic analysis and speech synthesis. Our study focuses on text analysis and phonetic analysis of under-resourced Malay dialect. In this thesis, a framework of under-resourced Malay dialect TTS system has been proposed. Besides, we also propose approaches to bootstrap Malay dialect models using Standard Malay and multilingual resources for developing Malay dialects TTS systems. In our proposed framework, a semi-supervised approach to translate Standard Malay text corpus to Malay dialects text corpus has been proposed to solve the problem of limited written text of Malay dialects. In this problem, we propose a word and phrase alignment to obtain vocabularies and translation rules of Malay dialects for translation task. The results show that the precision and recall values are above 95%. Next, a new normalisation algorithm has been proposed for standardising the orthography of Malay dialects text. For building a TTS system, it requires a standardised written text.

The results show that the vocabularies of dialect words are successfully standardised based on Standard Malay orthography with high accuracy. To find the sets of Malay dialects phonemes, an automatic phoneme identification approach using multilingual resources has been proposed. The results show that there are forty-seven phonemes of Kelantan dialect and thirty-four phonemes of Sarawak dialect found. With the information of phoneme and grapheme-phoneme mapping, a rule-based Malay dialects grapheme to phoneme (G2P) approach was then proposed to generate Malay dialects pronunciation dictionaries. The experiment results show that phoneme accuracy and word accuracy achieve above 90%. Finally, we evaluate our proposed framework of the Malay dialects TTS systems through perception test. The result shows that the quality of the Malay dialects speech synthesised through hidden Markov model (HMM) speech synthesis system using our proposed framework shows that from the scale of 1 (bad) to 5 (excellent), the average rating given is scale 4, which is "good".

# CHAPTER 1

# INTRODUCTION

## 1.1 Introduction

Dialect text-to-speech (TTS) system is an interesting area of research. It is an artificial speech production system that converts text to speech in dialect. A dialect is defined by linguists as a variety of a language that is distinguished from other varieties of the same language by its pronunciation, grammar, vocabulary, discourse conventions, and other linguistic features (Solano-Flores, 2006). It is important to learn dialects as they are still flourished and widely used in many areas especially for unofficial matters. Speakers with the same dialect share the same origin, culture and social group. Often, to get acceptance into the group, a speaker must speak the same dialect. Some of the larger dialectal divisions in the English-speaking world are British English, American English and Australian English.

The quality of TTS systems is improving steadily allowing it to be equipped and embedded in many tools, for instance in reading, communication, education, and entertainment. TTS system can be the reading aids for the blind. Its first real use was in reading systems for the blind, where a system would read some text from a book and convert it into speech. The first commercial TTS application was probably the Kurzweil reading machine for the blind introduced by Raymond Kurzweil in the late 1970's. It consisted of an optical scanner and text recognition software and was capable of producing quite intelligible speech from written multi-font text (Klatt, 1987). The dialect TTS system can be applied or beneficial in reading. In a mobile phone, it can be used to read out short text messages (SMS) when someone texts in dialect. It is very convenient when one is driving as it is eyes-free while getting the content of the messages.

TTS system is very important and useful especially for people with communication difficulties such as deafened and vocally handicapped. People who are born deaf cannot learn to speak properly and people with hearing difficulties have usually speaking difficulties. When individuals lose the ability to produce their own speech for neurological or other reasons, alternative augmentative communication (AAC) devices (Doyle & Phillips, 2001) can be used. An AAC device with a speech synthesis capability is referred to as a ‘‘Voice Output Communication Aid’’ or VOCA. A well-known scientist who utilises a VOCA device in his daily life is Professor Stephen Hawking due to the Motor Neurone Disease (MND). People with such speech disorders lose a functional means of communication where they may withdraw from social interaction, and even from interaction with their own family. Therefore, a dialect speech synthesiser can be served as a tool to communicate with local people who do not understand sign language.

TTS system can be applied in education. It is used to help with spelling and pronunciation when a person is learning a new language and can create opportunities to learn when no human teacher is available (Lemmetty, 1999). It may also be very helpful for those who are impaired to read as they can slow down the speed of spoken language from speech synthesisers and train their articulation with a novel sound sequence. Besides, a speech synthesiser is a helpful aid to proofreading. It helps us to go through a document aurally, which can be a compliment, and improves upon visual checking for grammar and style issues because it is more effective in detecting grammatical and stylistic problems when listening than reading (Klatt, 1987). Spelling mistakes may also become easier to detect. Hence, dialect TTS system can be a teacher for teaching a particular dialect.

Other than these usages, there are still a lot of applications equipped with TTS technologies, for example, in entertainment productions such as games and animations. TTS system is used to synthesise the sound of narrators and NPCs in video games or players' communication in multiplayer role-playing games. In Avatar animations, speaking avatars use the synthesised speech sound. Therefore, TTS system is very useful and important in our daily life. Of course not limited to these, the areas of application for TTS are still expanding. The greatest improvements when it comes to speech synthesis were during the last 10 years, synthetic voices are becoming very natural. At the same time, the time is clearly not yet ripe for some other areas of TTS, for example in low resources dialectal synthesis (Keller & Zellner-Keller, 2000), and in emotional speech synthesis.

There are several Malay dialects can be found in Malaysia which can be grouped according to the state such as Kelantan dialect from the state of Kelantan, Terengganu dialect from the state of Terengganu and Kedah dialect from the state of Kedah. Building a Malay dialect TTS system is very useful. From a linguistic viewpoint, it will help us to understand and appreciate the interesting differences of dialects, which are important to help preserve the dialect and culture in it. Secondly, a TTS system will be useful for people who like to learn a particular dialect or it can be used in places that require this facility. Furthermore, the system also has vast potential to be applied in many areas that require speech generation or synthesis application.

## 1.2 Problem Statements

The challenges in building Malay dialects TTS systems are that the resources for Malay dialects are limited and most of the conventional approaches for modelling use a lot of resources. The resources required for building Malay dialects TTS systems includes Malay dialects pronunciation dictionaries, Malay dialects written text and Malay dialects speech corpora. However, there are some problems in acquiring the Malay dialects resources as there is inexistence of Malay dialects pronunciation dictionaries, quasi-unknown of the sets of Malay dialects phonemes, no standard orthography for Malay dialects and limited Malay dialects written texts.

**Problem 1: Inexistence of Malay dialects pronunciation dictionaries**

A pronunciation dictionary contains words and their corresponding pronunciation. The pronunciation of a word is determined by a sequence of phone or other speech sounds. It is important for building TTS system as conversion of text to a phonemic specification of pronunciation is a necessary step in all current approaches to TTS synthesis. For example, the word *ibu 'mother'* in Standard Malay is converted to a pronunciation of /i b u/ (Tan & Ranaivo-Malacon, 2009). However, there is inexistence of Malay dialects pronunciation dictionaries. Most of the native dialect speakers learned Malay dialects through their daily conversation. Therefore, Malay dialects pronunciation dictionaries need to be generated for building Malay dialects TTS systems.

**Problem 2: Quasi-unknown of the sets of Malay dialects phonemes**

A phoneme is the smallest contrastive unit in the sound system of a language. Each language or dialect might have different phoneme set. For example, there are thirty-six phonemes in Malay (Maris, 1979) and forty-four phonemes in English (Dawn, 2002). The set of phonemes for a specific language or dialect is important in developing pronunciation dictionary. Since there is quasi-unknown of the sets of Malay dialects phonemes, the sets of phonemes for Malay dialects need to be found out so that Malay dialects pronunciation dictionaries can be generated. To determine the sets of Malay dialects phonemes, some acoustic analysis (Rezaei & Salehi, 2006) can be carried out for analysing the recorded speech sound. It can be done manually but this approach is time-consuming. Therefore, an automatic way to determine the sets of phonemes used in Malay dialects will be very useful.

**Problem 3: No standard orthography for Malay dialects**

Of the many languages in the world, most actually are only spoken and do not have a writing system. Even for many of the languages that do have writing systems, the orthography is poorly standardised (Palkar et al., 2012). The single greatest challenge to all aspects of building under-resourced TTS system is the orthography issue. The resource has been the lack of any standardised writing system or spelling norms and consequent writing of one word multiple ways, and different words the same way (Kathol et al., 2005). One of the problems when collecting written text in Malay dialects is that there is no standard orthography. Native dialect speakers write based on their transcribing norm. Thus, various types of spelling for a certain word might occur. For example, the word ‘rumah’ in Standard Malay (English: house) might be written as ‘ghumoh’ or ‘rumoh’ in

Kelantan dialect. This may lead to ambiguity of written text as a word will have more than one spelling. Therefore, normalising written text for Malay dialects is one of the crucial steps for building Malay dialects TTS systems.

**Problem 4: Limited Malay dialects written texts**

Building a TTS system usually requires training data consisting of a speech corpus with corresponding transcripts. However, for the languages that are not written down in a standard manner, one can only find speech corpora. It may seem futile to build a TTS system when the language at hand does not have a text form. Indeed, if there is no text at training time, there would not be text at test time, and then one might wonder why we need a TTS system at all (Sitaram et al., 2013). Yet, the amount of written text in Malay dialects can be found is very limited as there are not many dialect speakers. Hence, obtaining a sufficient amount of Malay dialects text corpus is required.

## 1.3 Research Questions

The main research question in this study is "How to bootstrap under-resourced dialect models using existing resources to produce a good quality TTS system?"

In summary, the specific research questions are as follows:

1. How to create a pronunciation dictionary for a Malay dialect?
2. How to determine automatically the set of phonemes of a Malay dialect?
3. How to normalise the orthography of Malay dialects?
4. How to acquire written texts of Malay dialects?

## 1.4 Research Objectives

The objectives of the research are:

1. To propose a framework for under-resourced TTS system.
2. To develop Malay dialects pronunciation dictionaries using knowledge-based approach through grapheme to phoneme (G2P) conversion rules.
3. To propose an approach to automatically determine the set of phonemes in a dialect using multilingual resources.
4. To normalise and standardise Malay dialects text.
5. To propose a way to obtain Malay dialects text corpus from existing Standard Malay text corpus.

## 1.5 Importance of the Research

Owing to a recent improvement in speech quality, TTS system is now used in various fields. For example, it can be used for information interchange between the user with weak eyesight and the computer (Aida–Zade et al., 2013), educational technology purpose (Black, 2007), car navigation system, video game machines (Shiga & Kawai, 2012) and others. Research in TTS for a language with no standard orthography is important in the development of a speech to speech translation system from a language to another language. For example, there is a speech-to-speech translation of video lectures from English into Konkani (Sunayana et al., 2013).

In this study, Malay dialects are focused. It is important to be studied as it helps to define different regions in Malaysia and also one's identification. There are some different Malay dialects that can be grouped by states in Malaysia, for example, Perak dialect from the state of Perak and Kedah dialect from Kedah. By observing the type of Malay dialects that one used to communicate with, their living cultures can be studied. It is important to learn Malay dialects as they are still flourished and widely used. Speakers with the same dialect share the same origin, culture and social group. However, younger generation nowadays may not be aware of their own Malay dialect. It will cause some of the Malay dialects are on the brink of extinction and will completely disappear in the future. Therefore, building Malay dialects TTS systems will help us to understand and appreciate the interesting differences of Malay dialects in Malaysia in order to help preserve the dialect and culture in it. The Malay dialects TTS systems built will be useful for people who like to learn a particular dialect. Besides, the system is useful in some different places that require TTS technologies such as in local animation film. It can be also a useful tool for communicating with the local. Last but not least, the proposed system has vast potential to be equipped with many other applications.

## 1.6 Research Scope

This research focuses on resource acquisition, text analysis and phonetic analysis modules in building Malay dialects TTS systems. Refer Figure 3.1. Resource acquisition is challenging especially in collecting Malay dialects written text corpus as there is a lack of written text in Malay dialect. Text analysis includes text normalisation and linguistic analysis of Malay dialect. The written text in Malay dialects available is in unstandardised form. Therefore, normalisation of the Malay dialects text is important to standardise the orthography in Malay dialect. Besides, linguistic analysis is essential for analysing the grammatical structures and unique vocabularies used in Malay dialects. Furthermore, a phonetic analysis is necessary for converting Malay dialects text into phonetic alphabets. As there is no formal pronunciation dictionary for Malay dialect, grapheme to phoneme

(G2P) conversion can be conducted. However, there is quasi-unknown of the sets of Malay dialects phonemes. Hence, the sets of Malay dialects phonemes need be found out. Then, the pronunciation of the Malay dialects words can be developed using grapheme to phoneme (G2P) conversion.

Since Malay dialects are under-resourced languages, HMM speech synthesis is chosen as the resources collected for Malay dialects are limited. The speech can be synthesised with voice characteristics of a target speaker by adapting HMMs with only a small number of utterances. Besides, more variations are allowable, for example, the original voice can be converted into another voice. Furthermore, less memory is needed to store the parameters of the models than to store the data itself. HMM speech synthesis can be built quickly and produces good quality synthetic voice where it can synthesise a smooth and natural sounding speech. Malay dialect speech will be synthesised from Standard Malay but not from the dialect itself.

In this thesis, Kelantan dialect and Sarawak dialect are chosen. Kelantan dialect is one of the Malay dialects from the north-east of Peninsular Malaysia in ‘schwa-varieties’ group where Sarawak dialect is from East Malaysia in ‘a-varieties’ group. Although there are many Malay dialects in Malaysia, this research may not include all Malay dialects due to time constraint. However, the proposed approaches could be applied to most of the Malay dialects

## 1.7 Research Methodology

A framework for under-resourced Malay dialect TTS system is proposed. In our framework, we solve the problems of limited language resources in Malay dialects of inexistence of pronunciation dictionaries, quasi-unknown of the sets of phonemes, no standard orthography and limited written text. Besides, new approaches are proposed to solve problems in building a Malay dialect TTS system. The following steps outline the methodology employed by the research.

### A. Creating Parallel Texts in Standard Malay and Malay Dialects

A parallel text is required to obtain Malay dialects translation rules and vocabularies through alignment so that translation task can be conducted from Standard Malay (source language) text to Malay dialects (target language) text as there is a large amount of existing Standard Malay text corpus. To create parallel texts of Standard Malay and Malay dialect, conversations that transcribed and translated by dialect speakers need to be collected.

### B. Word and Phrase Alignment

A word and phrase alignment approach on the parallel texts of Standard Malay and Malay dialects is proposed which take into consideration of the similar writing between the source language and target language. It is to obtain Malay dialects translation rules and vocabularies for translation task.

### C. Text Normalisation

A new normalisation approach for standardising the orthography of Malay dialects is proposed in order to standardise the written form of Malay dialects. Our proposed text normalisation focusses on normalising and standardising the vocabularies that are perceptually and semantically similar in Standard Malay and Malay dialect; and unique vocabularies that do not exist in Standard Malay.

**D. Malay Dialect Text Generation**

A semi-automatic approach is used to translate Standard Malay text corpus to Malay dialects text corpus.

**E. Automatic Phoneme Identification for Dialects**

An automatic phoneme identification approach using multilingual resources is proposed to find out the sets of the phonemes of Malay dialects. Only the number of unique phonemes used in Malay dialects need to be found out but not the actual type of it because every phoneme is represented by using a symbol or alphabet.

**F. Knowledge-based Grapheme to Phoneme**

A rule-based Malay dialects grapheme to phoneme (G2P) approach is proposed to generate Malay dialects pronunciation dictionaries.

**G. Read Speech Transcript Construction**

Read speech transcript is constructed for recording so that speech synthesis corpus can be obtained which used to train a dialect acoustic model for building a Malay dialect TTS system. The read speech transcript is selected from a large amount of translated Malay dialect text corpus with phonetically-rich text.

**H. Malay Dialect HMM-based Speech Synthesis System**

Malay dialects speech is synthesised using HMM-based speech synthesis system. HMM speech synthesis is chosen for building Malay dialect TTS system as the speech can be synthesised using a small amount of speech data. Besides, there is much less development time and also much lower development cost.

## 1.8 Contributions

Our contributions in this research are mainly focused on resources acquisition, text analysis and phonetic analysis in Malay dialects TTS systems. A framework for under-resourced Malay dialect TTS system was proposed. In our framework, the problems of limited language resources in Malay dialects were solved which include inexistence of Malay dialects pronunciation dictionaries, quasi-unknown of the sets of Malay dialects phonemes, no standard orthography in Malay dialects and limited written text of Malay dialects.

**A. Resources Acquisition**

i. Conversation recording was acquired on Kelantan dialect and Sarawak dialect.

ii. Parallel texts of Kelantan dialect/Standard Malay and Sarawak dialect/Standard Malay were created.

iii. A large amount of translated text corpus in Kelantan dialect and Sarawak dialect without manual checking were obtained.

iv. Malay dialects read speech transcript in Kelantan dialect and Sarawak dialect were constructed.

v. Speech synthesis corpus in Kelantan dialect and Sarawak dialect were obtained.

**B. Text Analysis**

i. A word and phrase alignment approach was proposed by taking into consideration of the similar writing between the Standard Malay and Malay dialects to find out unique vocabularies and translation rules of Malay dialects for translation task.

ii. A text normalisation approach was proposed by considering the vocabularies that are perceptually and semantically similar in Standard Malay and Malay dialect; and unique vocabularies that do not exist in Standard Malay.

**C. Phonetic Analysis**

i. An automatic phoneme identification approach was proposed using multilingual phoneme recognizer to determine the set of phonemes of a Malay dialect.

ii. A knowledge-based grapheme-to-phoneme (G2P) conversion was proposed to develop pronunciation dictionary of Malay dialects.

## 1.9 Organisation of Thesis

The following chapters of the thesis is organised as follows:

In Chapter 2, research background of Malay dialects and synthesis techniques will be reviewed.

In Chapter 3, a literature review of text analysis and phonetic analysis for TTS system will be discussed. The chapter will start with the discussion of the general architecture of TTS system, followed by text corpus construction, text normalisation, parallel text alignment, phoneme identification, grapheme to phoneme (G2P), under-resourced speech synthesis, the current status of research for Malay speech synthesis and finally the summary of the literature review.

In Chapter 4, the proposed methodology of the research is included. The work of this research that bootstrap under-resourced dialect models using existing resources to build Malay dialect TTS system is explained in details in this chapter.

Chapter 5 discussed some experiments conducted in this research. The result of each experiment is analysed in details.

In Chapter 6, the discussion on each proposed approach is presented.

In Chapter 7, it presents the summary of the research outcomes. It revisits and reconfirms that the research objectives are met and the contributions are achieved. Future work is also included.

Figure 1.1 illustrates research problems, research questions and research objectives mapping.

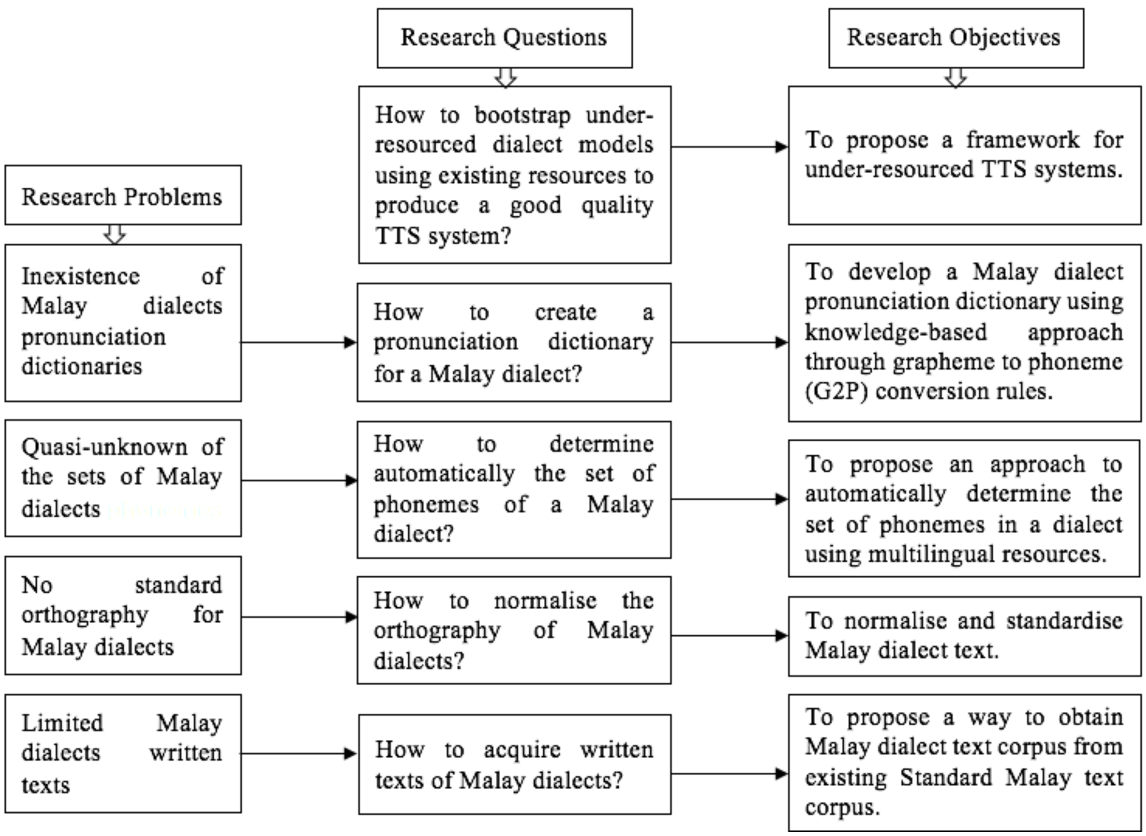


**Figure 1. 1 Research Problems, Research Questions and Research Objectives Mapping**

# CHAPTER 2

# RESEARCH BACKGROUND

## 2.1 Malay Dialects

Malay is a language from the Austronesian family. It is the official language of Malaysia, Indonesia, Singapore, and Brunei. However, Malay spoken in different countries and even within a country itself might vary in terms of pronunciation and vocabulary from one place to another where dialect occurred. Malay dialect is very distinctive as every single pronunciation of the words can bring various meanings. Besides, Malay dialects do not have written form and thus native speakers learn from their daily conversation only. In Malaysia, the formal Malay language recognised and used is known as Standard Malay. Standard Malay is originated from Johor, Riau dialect variety. The prominence of Johor, Riau dialect is due to the influence and importance of empire in the 19th century. There are several Malay dialects found in Malaysia that can be grouped according to the geographical distribution (Colins, 1989). Malay dialects in Peninsular Malaysia are classified into seven groups, the North-Western group comprising Kedah, Perlis, Penang and North Perak dialects; the North-Eastern group, that is, the Kelantan dialect; the Eastern group, that is the Terengganu dialect; the Southern group comprising Johor, Melaka, Selangor and Perak (Southern); the Negeri Sembilan group; the Pahang dialect as a group by itself and not as a member of Southern group and the Perak dialect, the latter of which covers the area of Central Perak (Asmah, 1991). Figure 2.1 shows the Malay dialects in Peninsular Malaysia.

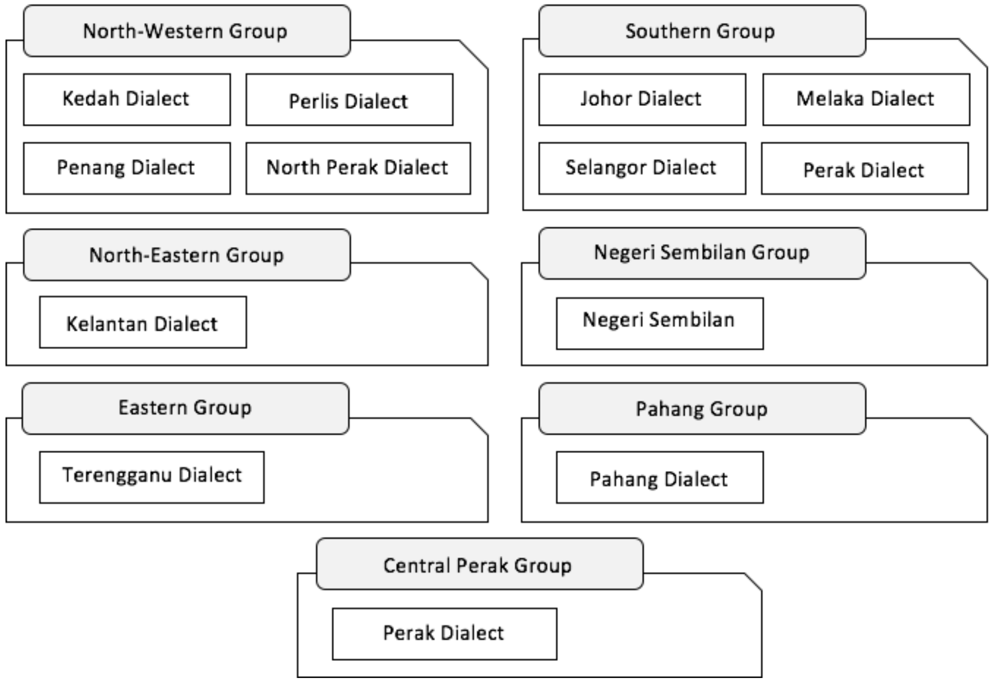


**Figure 2. 1 Malay Dialects in Peninsular Malaysia**

Each group may be further classified to different subdialects according to different areas. For example, Malay dialects spoken in Perak (northern state of Malaysia) can be classified according to five areas. The northern part of Perak speaks Petani and Kedah dialect, the southern part speaks Selangor dialect, slightly to the east part speaks Rawa dialect, while the area around the middle of Perak around Parit and Kuala Kangsar speaks Perak dialect (Ahmad, 1991). In Malaysia this has given rise to two groups of standard varieties, which can be termed the ‘a-varieties’ and the ‘schwa varieties’ as they differ in the realisation of word-final orthographic ⟨a⟩ (pronounced as either /a/ or / ə/) in addition to a range of other features (Omar, 1977). Figure 2.2 shows the Malay dialects spoken by Malays and non-Malays in Peninsular Malaysia.


**Figure 2. 2 Malay Dialects Spoken by Malays and Non-Malays in Peninsular Malaysia** (Omar, 2011)

The difference between each Malay dialect is systematic in term of pronunciation, vocabulary and grammatical structure. The number and types of phonemes used for each Malay dialect would be slightly different as each of them has their own unique pronunciation modelling. For example, the pronunciation of the word *kuning 'yellow'* in Standard Malay will be /k u n ĩ/ in Kelantan dialect, /k u n i n/ in Perak dialect and /k u n e n/ in Kedah dialect (Ajid, 1977). Malay dialect differs not only in the pronunciation, but it might also vary in term of vocabulary. There will be some additional vocabularies used in each Malay dialect compare to Standard Malay. For example, in Kedah, the word for *kamu 'you'* is *hang*, *air 'water'* is *ayak* and *sen* '*cent'* is called *kupang*. In term of grammar, they are almost similar. There are not many words that a dialect differs from the Standard Malay where around eighty percent of the words are shared by Standard Malay and Malay dialects. Nevertheless, there is no formal written form for Malay dialects. Besides, Malay dialect speakers learn their dialect through their daily conversation without a formal pronunciation dictionary.

There are some phonology and vocabulary of Malay dialects analysis in previous studies, for example, Kedah dialect, Perak dialect, Kelantan dialect and Sarawak dialect (Asmah, 1977; Asmah, 1988). Abdul (2006) presented the synthesis of Kelantan dialect that includes phonology, vocabularies and sentence structures. The phonology of Sarawak dialect also have been discussed (Johari, 1989). An example of a list of phonemes in Kelantan dialect and Sarawak dialect are shown in Table 2.1 and Table 2.2.

**Table 2. 1 List of Phonemes in Kelantan Dialect**

| **Vowels** | **Consonants** | |
|---|---|---|
| /a/ | /p/ | /m/ |
| /e/ | /b/ | /n/ |
| /i/ | /t/ | /ŋ/ |
| /o/ | /d/ | /ɲ/ |
| /u/ | /k/ | /w/ |
| /ə/ | /g/ | /j/ |
| /ɔ/ | /s/ | /dʒ/ |
| /ã/ | /h/ | /l/ |
| /ẽ/ | /f/ | /r/ |
| /ĩ/ | /v/ | /z/ |
| /ɛ̃/ | /ʃ/ | /ʔ/ |
| /ũ/ | /x/ | Double Consonants |
| /õ/ | /ɣ/ | |
| /ɔ̃/ | /tʃ/ | |

**Table 2. 2 List of Phonemes in Sarawak Dialect**

| Vowels | Consonants | | Diphthongs |
|---|---|---|---|
| /a/ | /p/ | /tʃ/ | /aw/ |
| /e/ | /b/ | /m/ | /aj/ |
| /i/ | /t/ | /n/ | /oj/ |
| /o/ | /d/ | /ŋ/ | |
| /u/ | /k/ | /ɲ/ | |
| /ə/ | /g/ | /w/ | |
| | /s/ | /j/ | |
| | /h/ | /dʒ/ | |
| | /f/ | /l/ | |
| | /v/ | /r/ | |
| | /ʃ/ | /z/ | |
| | /x/ | /ʔ/ | |
| | /ɣ/ | | |

However, the vocabularies and the pronunciation used can be varied during the lifespan of an individual. Besides, it can be changed due to the dialect of different cohorts of individuals living within a speech community.

## 2.2 Synthesis Techniques

The first synthetic speech was produced since the late 18th century (Trouvain & Brackhane, 2011). Kempelen's speaking machine was probably the first ever functioning mechanical speech synthesiser that was able to generate short utterances. Figure 2.3 shows the main components of von Kempelen's speaking machine. The machine was built in wood and leather.

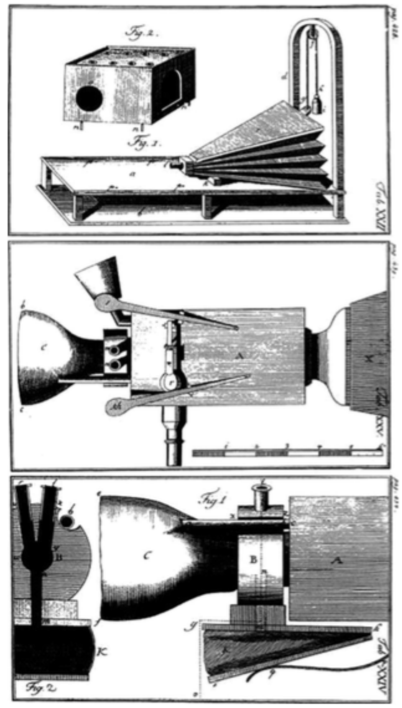

**Figure 2. 3 The Main Components of von Kempelen's Speaking Machine** (Flanagan, 1972)

From the figure, it shows that a bellow is representing the lungs, a rubber tube for the mouth, and a wooden extension being the nose. By means of two levers controlling the resonance characteristics, a complete set of a language's sounds could be produced. Von Kempelen successfully produced whole words and sentences in languages such as Latin, French and Italian (Ritter, 1791). In the early 20th century, electronic synthesisers were created. The electronic device that can be considered as a speech synthesiser was Voice Operating Demonstrator (VODER) introduced by Homer Dudley in New York World's Fair 1939 (Flanagan, 1972; Klatt, 1987). VODER was inspired by Voice Coder (VOCODER) developed at Bell Laboratories in the mid-thirties. The original VOCODER was a device for analysing speech into slowly varying acoustic parameters which could then drive a synthesiser to reconstruct the approximation of the original speech signal. The VODER consisted of wrist bar to select a voicing or noise source and a foot pedal for controlling the fundamental frequency. The source signal was routed through ten bandpass filters whose output levels were controlled by fingers. It took considerable skill to play a sentence on the device. The speech quality and intelligibility were far from good but the potential for producing artificial speech was well demonstrated. Figure 2.4 shows the VODER speech synthesiser.

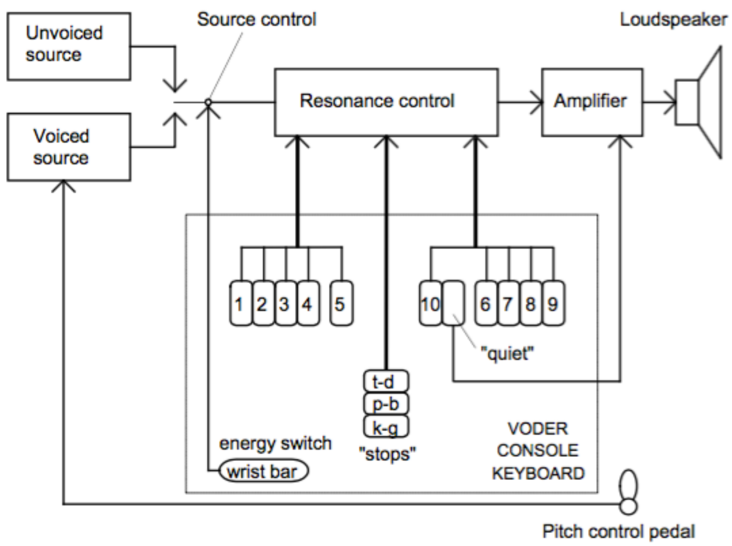


**Figure 2. 4 The VODER Speech Synthesiser** (Klatt, 1987)

### 2.2.1 Formant Synthesis

Several types of speech synthesis techniques were introduced over the years, for example, formant synthesis, concatenative synthesis, articulatory synthesis and HMM-based synthesis. The first formant synthesiser, PAT (Parametric Artificial Talker), was introduced by Walter Lawrence in 1953 (Klatt, 1987). A formant is a concentration of acoustic energy around a particular frequency in the speech wave. Roughly there is one formant in each 1000Hz band. Each formant corresponds to a resonance in the vocal tract. Figure 2.5 shows the example of formant.

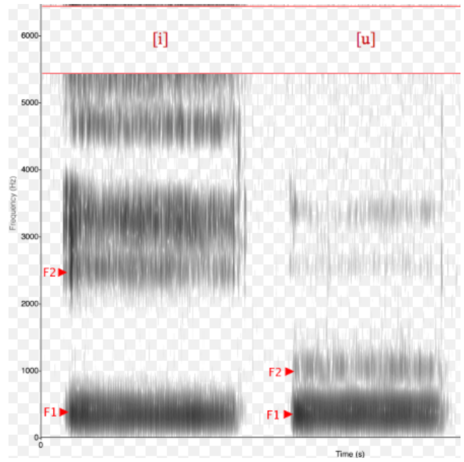


**Figure 2. 5 Example of Formant**

Formant synthesis system is the oldest method for speech synthesis, and it dominated the synthesis implementations for a long time. It also known as rule-based synthesis, creates the acoustic speech data entirely through rules on the acoustic correlates of the various speech sounds. It can well simulate human processes of speech generation in a frequency domain, flexible control of acoustic features is possible (Klatt, 1980; Klatt & Klatt, 1990). No human speech recordings are involved at the run time, as a result, the synthetic voice sounds robotic (Balyan et al., 2013). Once a formant-based synthesiser is built with a relatively small amount of speech data, it may be possible to generate high-quality synthetic speech of various speaking styles and emotions. The vocal tract is simulated by a sequence of second order filters in cascade while a parallel structure is mostly used for the synthesis of consonants (O'Saughnessy, 1987). Figure 2.6 shows the basic structure of cascade formant synthesiser.

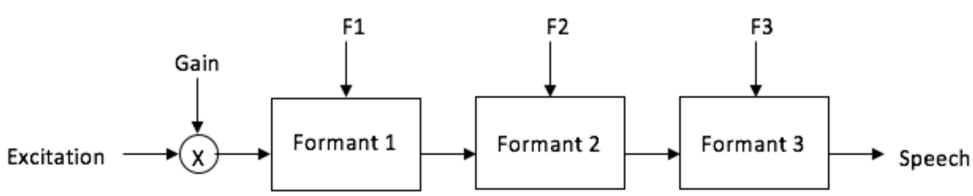


**Figure 2. 6 Basic Structure of Cascade Formant Synthesiser**

### 2.2.2 Articulatory Synthesis

The articulatory synthesiser was introduced in 1958 by George Rosen at the Massachusetts Institute of Technology, M.I.T. (Klatt, 1987). Articulatory synthesis is by far the most complicated in regard to the model structure and the computational burden. Articulatory synthesis is the production of speech sounds using a model of the vocal tract, which directly or indirectly simulates the movements of the speech articulators (Wu & Hsieh, 2000). In the articulatory model, the vocal tract is divided into many small sections and the corresponding cross-sectional areas are used as parameters to represent the vocal tract characteristics. Figure 2.7 shows the example of direction variation for parameters of the articulatory model.

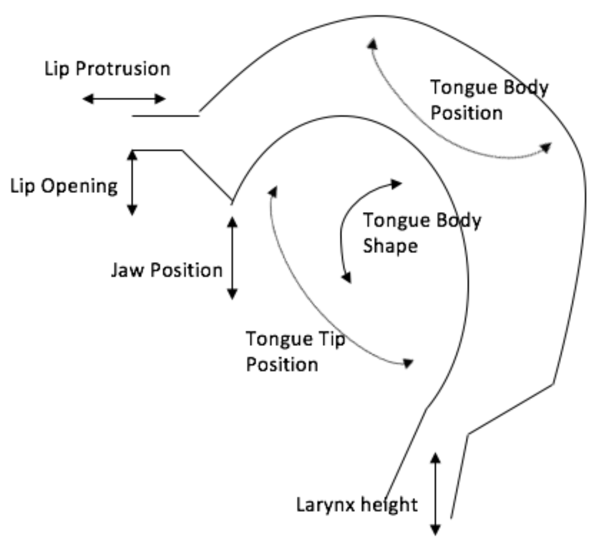


**Figure 2. 7 Direction of Variation for Parameters of Articulatory Model**

Articulatory synthesis systems contain physical models of both the human vocal tract and the physiology of the vocal cords. It is common to use a set of area functions to model the variation of the cross-sectional area of the vocal tract between the larynx and the lips. In the acoustic model, each cross-sectional area is approximated by an electrical analogue transmission line. The area functions must change with time in order to simulate the movement of the vocal tract. Each sound is designated in terms of a target configuration and the movement of the vocal tract is specified by a separate fast or slow motion of the articulators. The principle is thus similar to the one that has been seen within the acoustic tube model. The articulatory model involves a large number of control parameters that are used for the very detailed adjustment of the position of lips and tongue, the lung pressure, the tension of vocal cords, and so on. The method of articulatory synthesis is not among the best for the quality of produced speech sounds. The synthesis produces intelligible synthetic speech. However, its output is still far from the natural sound (Balyan et al., 2013; Tatham & Morton, 2005). It represents an attempt to do articulatory synthesis with limited control parameters, a relatively small set of which are likely to have an effect on the acoustic output for both vowels and consonants (sonorants and obstruents).

### 2.2.3 Concatenative Synthesis

Concatenative synthesis is one of the most popular and successful techniques until today because of the high quality of synthetic speech that it can be produced. It uses a large amount of speech data, segmented into units, and a unit selection algorithm that finds the sequence of units that match best the sound or phrase to be synthesised (Tatham & Morton, 2005). Figure 2.8 shows the concatenative synthesiser.

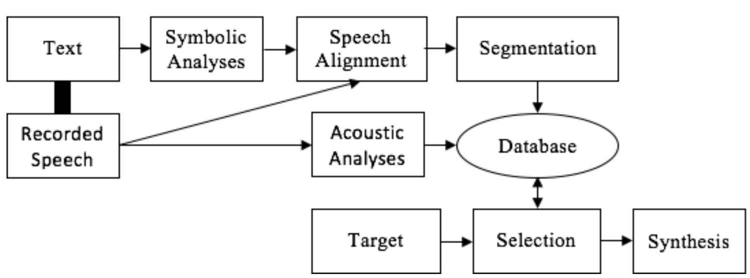


**Figure 2. 8 Concatenative Synthesiser**