UNIVERSITI TEKNIKAL MALAYSIA MELAKA

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

DEVELOPMENT OF AN INTERACTIVE SYSTEM USING VOICE PROCESSOR

This report submitted in accordance with requirement of the Universiti Teknikal Malaysia Melaka (UTeM) for the Bachelor Degree of Engineering Technology (Industrial Power) (Hons.)

by

IZZAT AMINUDIN BIN MOHAMAD SALJI
B071210370
900802-08-5709

FACULTY OF ENGINEERING TECHNOLOGY
2015

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

BORANG PENGESAHAN STATUS LAPORAN PROJEK SARJANA MUDA

TAJUK: **Development of An Interactive System Using Voice Processor**

SESI PENGAJIAN: **2015/16 Semester 1**

Saya **IZZAT AMINUDIN BIN MOHAMAD SALJI**

mengaku membenarkan Laporan PSM ini disimpan di Perpustakaan Universiti Teknikal Malaysia Melaka (UTeM) dengan syarat-syarat kegunaan seperti berikut:

1. Laporan PSM adalah hak milik Universiti Teknikal Malaysia Melaka dan penulis.
2. Perpustakaan Universiti Teknikal Malaysia Melaka dibenarkan membuat salinan untuk tujuan pengajian sahaja dengan izin penulis.
3. Perpustakaan dibenarkan membuat salinan laporan PSM ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. **Sila tandakan (✓)

☐	SULIT	(Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan)
☐	TERHAD	(Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia sebagaimana yang termaktub dalam AKTA RAHSIA RASMI 1972)
☒	TIDAK TERHAD	

Disahkan oleh:

Alamat Tetap:

LOT 7010 Matang Pasir

Alor Pongsu 34300 Bagan Serai

Perak Darul Ridzuan

Tarikh: 12/1/2016

Cop Rasmi:

AHMAD IDIL BIN ABDUL RAHMAN
Pensyarah Kanan
Jabatan Teknologi Kejuruteraan Elektrik
Fakulti Teknologi Kejuruteraan
Universiti Teknikal Malaysia Melaka

** Jika Laporan PSM ini SULIT atau TERHAD, sila lampirkan surat daripada pihak berkuasa/organisasi berkenaan dengan menyatakan sekali sebab dan tempoh laporan PSM ini perlu dikelaskan sebagai SULIT atau TERHAD.

DECLARATION

I hereby, declared this report entitled "Development of An Interactive System Using Voice Processor" is the results of my own research except as cited in references.

Signature : .....

Name : Izzat Aminudin

Date : 12/1/2016

APPROVAL

This report is submitted to the Faculty of Engineering Technology of UTeM as a partial fulfillment of the requirements for the degree of Bachelor of Engineering Technology (Industrial Power) (Hons.). The member of the supervisory is as follow:

.....

(En. Ahmad Idil Bin Abdul Rahman)

ABSTRACT

This report presents an interactive system using voice processor to control the speech synthesis and speech recognition. The proper process is very important and need to be taken seriously in this project to obtain better accuracy and output. Unfortunately, ordinary attempt and mistake strategy is tedious and also high cost. The purpose for this research is to develop an interactive system using voice processor and simulation modelling of a speech synthesis program. Multiple types of controller will be study to choose what best for develop the voice processor system. The result from this research is useful to be implemented in our daily life to help much kind of people.

ABSTRAK

Laporan ini membentangkan satu sistem interaktif menggunakan pemproses suara untuk mengawal sintesis pertuturan dan pengecaman pertuturan. Proses yang betul adalah sangat penting dan perlu diberi perhatian serius dalam projek ini untuk mendapatkan ketepatan keluaran yang lebih baik. Malangnya, dengan mencuba *kaedah konvensional memakan masa serta kos yang tinggi. Tujuan kajian ini adalah* untuk membangunkan satu sistem interaktif menggunakan pemproses suara dan pemodelan simulasi program sintesis pertuturan. Jenis gandaan pengawal akan mengkaji untuk memilih apa yang terbaik untuk membangunkan sistem pemproses suara. Hasil daripada kajian ini amat berguna untuk dilaksanakan dalam kehidupan seharian kita untuk membantu banyak jenis orang.

DEDICATIONS

To my beloved parents

Encik Ahmad Idil Bin Abdul Rahman

Lectures of Faculty of Engineering Technology

Friends

ACKNOWLEDGMENTS

I am thankful and want to express my genuine appreciation to my supervisor Ahmad Idil Bin Abdul Rahman for his direction, persistent consolation and constant support during in the making of this research. I really appreciate his guidance from the earliest starting point to the end until I enabled to develop an understanding of this research completely. I also sincerely thank him for the time spent showing and redressing my oversights.

I recognize my sincere indebtedness and gratitude to my parents for their love, support and sacrifice throughout my life. Their sacrifice had motivated me from the day I figured how to read and write until what I have turned out to be currently. I can't locate the fitting words that could properly describe my thankfulness for their commitment, support and confidence in my ability to accomplish my study.

Many thanks go to my whole class part for their excellent co-operation, inspirations and supports during this study. This four year involvement with every one of you will be remembered as sweetest memory for me.

Lastly I might want to thanks any individual which contributes to my final year project straightforwardly or by implication. I might want to recognize their comments and suggestions, which was crucial for the successful fulfilment of this study.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLE

CHAPTER 1
INTRODUCTION

1.0 Introduction

Language teaching is fairly troublesome and muddled procedure that requires watchful and persistent work. Instructors in the field of dialect showing dependably make a decent attempt to discover approaches to make dialect learning pleasant and appealing for the learners. Distinctive exercise, activities, games and fascinating stories assisted dialect instructors with achieving this point through numerous years despite everything they do.

Today, we have access too many programs that are currently used and tested by using speech synthesis and voice recognition. Text-to-speech technology is a one example of application that is common feature that has been use in educational purpose. Text-to-speech is the ability of a computer to produce spoken words. Computer speech can be produced either by "splicing" pre-recorded words together or, with much more difficulty, by having the computer produce the sounds that make up spoken words.

In other words, text-to-speech is the conversion of text to speech through special computer applications. Text-to-speech software is priceless for blind computer users as it enables them to "read" from the screen. This innovation was first introduced as Texas Instruments Speak and Spell handheld electronic learning aid in 1978.

To make use of this technology, there has been something that can be done other than language learning using text to speech and speech to text conversion. Nowadays, people are likely want to make their life easier, so by using voice recognition and speech synthesis, it can combine it with electrical appliances and control many type of electrical component using voice recognition like turn on the light.

1.1 Background

In computer science, speech recognition (SR) is the translation of spoken words into text. It is also known as "automatic speech recognition" (ASR), "computer speech recognition", or just "speech to text" (STT).

Some SR systems use "speaker-independent speech recognition" while others use "training" where an individual speaker reads sections of text into the SR system. These systems analyze the individual's particular voice and utilize it to adjust the recognition of that person's speech, resulting in a more accurate transcription. Systems that do not use training are called "speaker-independent" systems. Systems that use training are called "speaker-dependent" systems.

Speech synthesis is the artificial production of human. A computer system used for this purpose is called a speech synthesizer, and can be implemented in our products. A text-to-speech (TTS) system converts normal language text into speech. Different systems render symbolic linguistic representations like phonetic transcriptions into speech.

1.2 History

Each sort of dialect showing uses its own method to help learners. With the introduction of grammar translation method, the blackboard came into use in language classrooms. Later, it was replaced by overhead projector. At that point, computer software was used to provide students with drill and practice exercises.

The first use of computer by institutions related to teaching and learning coincide with the introduction of second-generation computers towards the 1950s. The large universities started to use computers for administrative process and student record keeping. At the same time computers are used for instructional teaching and research. PLATO (Programmed Logic for Automatic Teaching Operations), the very first project related to use of computers in educational research, began in 1960 at the University of Illinois to design a large computer-based system for instruction. The system included a mainframe machine supporting hundreds of terminals which have

high capacity comparing to that age. Many course in many disciplines were develop, design and delivered on that systems (Alessi & Trollip, 1985; Warschauer, 1996; Levy, 1997; Culley, 1992).

During the 1960s and 1970s, the use of computer-assisted instruction expanded in public schools with the introduction of the next generation of computers and microchips which were cheaper (Bullough & Beatty, 1991). In 1971, another important project, TICCIT (Time-shared, Interactive, Computer Controlled Information Television) was initiated at Brigham Young University (Levy, 1977). The system combined television technology with the computer to deliver instruction to the learners.

In the 1980s, microcomputers started to be adopted by the schools and new developments such as CD-ROM, speech-based software and interactive video appeared. Also experiments were done in the integration of the computers into the curriculum. In the 1990s and 200s, with the introduction of fast, affordable *processors, new software, wide-scale and fast access to the internet made computers* available in almost all public and private schools as well as homes personal and educational use.

Meanwhile, what went unnoticed was the 'text-to-speech' technology basically designed for the visually impaired people. Speech synthesis is the conversion of text to speech through special computer applications that often referred as Text To Speech software (TTS). Text-to-speech software is considered invaluable for the blind since it enables them to read from the computer screens.

1.3 Problem Statement

A few individuals may have physical inabilities that block their abilities to accomplish something even to turn on the light for instance. For these people, voice recognition is one of a few option information strategies to be investigated. Voice recognition may give a more productive method for controlling an electrical component that is less physically and cognitively taxing than other option information methods. However, someone may appear to be able to utilize the ability

to use the keyboard, but have subtler physical difficulties that make voice recognition a more attractive option.

In addition, speech synthesis has long been a vital assistive technology tool and its application around there is noteworthy and far reaching. It allows environmental barriers to be removed for people with a wide range of disabilities.

In response to this problem, this project purpose is to study and develop the systems using voice processor to control the electrical component and to make it more interactive to use.

1.4 Objectives

The objectives were refined and created to be more particular achievable. All things considered, these achievable objectives for this project are:

1. To make a system that can control electrical equipment using arduino uno
2. To build an android application that can be used with arduino uno
3. To create modern technology using voice processor with more interactive

1.5 Scope

Project scope is the limitation for each project that has been conduct. One of the scopes for this project is focused on an ideal combination of an android application with the connection of arduino uno microcontroller.

The bluetooth device will be used to connect between the android application and the arduino uno microcontroller. The android application need to be making and installed in an android smartphone as a voice to text process.

CHAPTER 2
LITERATURE REVIEW

2.0 Introduction

This chapter will provide the review from previous research that is related to this project. There are previous researchers on speech synthesis and speech *recognition that using different type of application and method to obtain the certain* system.

2.1 Text to Speech

Text-to-speech is a speech synthesis system is by definition a system, which produces synthetic speech. It is implicitly clear, that it involves some sort of input. What is not clear is the type of this input. If the input is plain text, which does not contain additional phonetic and/or phonological information, the system may be called a text-to-speech (TTS) system. A schematic of the text-to-speech process is shown in the figure 2.1 below. As shown, the synthesis starts from text input. Nowadays, *this may be plain text or marked-up text, for example HTML or* something similar like JSML (Java Synthesis Mark-up Language).

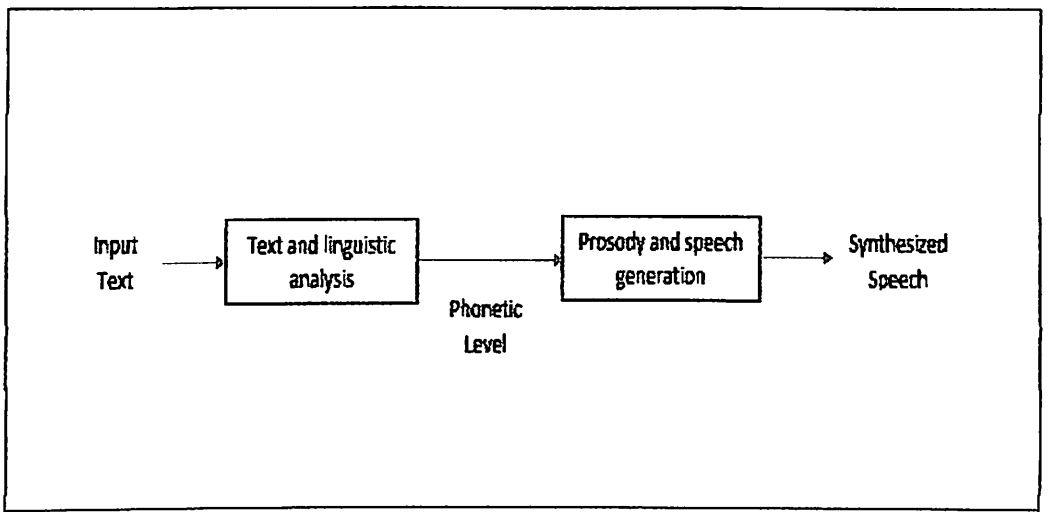

Figure 2.1: Text to Speech Process

2.2 HMM-based Speech Synthesis

In order to modelling speech synthesis system, several methods had been used in previous research. (Yan-You Chen, 2013). He has done analysis on speech variability compensation for expressive speech synthesis and purposes the suitable features that are used to generate the synthetic speech by the vocoder. For this purposes, he states that the variance of speech distribution is still rarely examined in conventional speech synthesis. The system output with the same result and leads to the over-stability problem are cause by the ignorance. However, the training data is stable and has a lower variance of distribution in nature, which means that the problem is not manifest in reading style or natural speech synthesis. Another problem is settled if the speech variability increases.

In this research, he also investigates a novel method on speech variability to overcome the drawback which suffers from the over-stability in conventional speech and the degree of perception is improved for expressive speech synthesis. The method can generate more natural speech parameters regarding the speech variability. In addition, this method also makes the synthesized speech regarding the time-variance to achieve the human-like expressive speech not only an attempt to counter the disadvantage over traditional speech synthesis in over-stability.

Meanwhile in system overview for this research, he has combined his proposed method with a traditional HMM-based speech synthesis system in a system block diagram in figure 2.2. It consists in two phases which are training phase and synthesis phase. For training phase, the spectrum and excitation parameters are extracted and modelled by the context dependent HMMs. In synthesis phase, spectrum and excitation parameters are generated from the context-dependent HMMs and the duration models for a given text which is converted into a context-dependent label sequence by text analyzer.

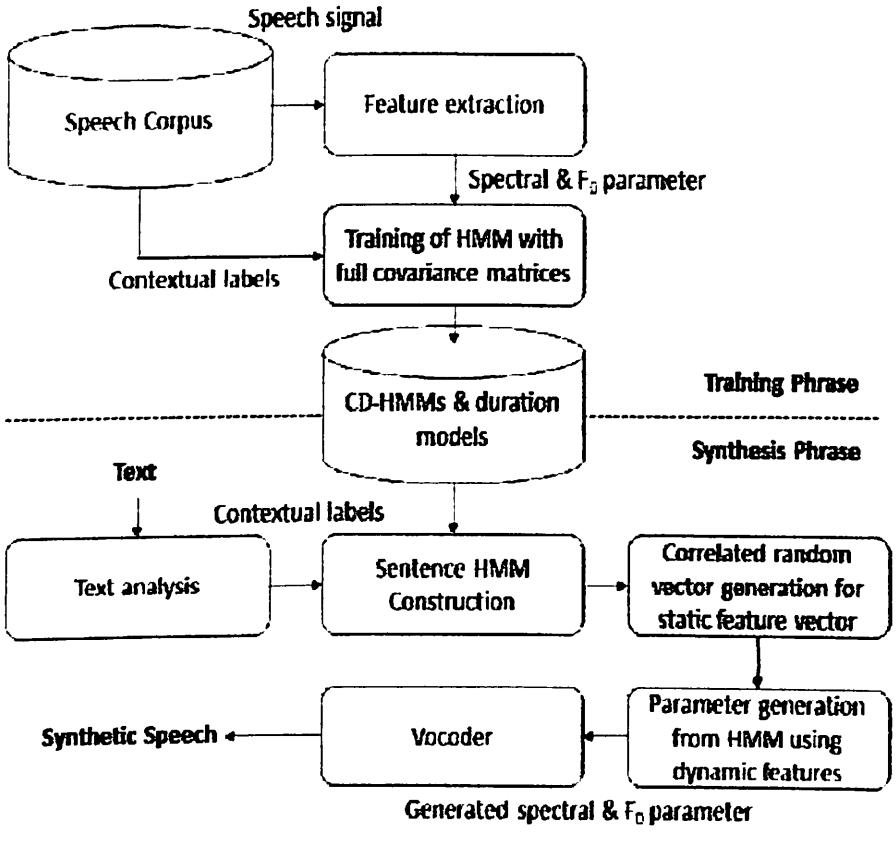

Figure 2.2: Overview of HMM-based Speech Synthesis

However, because of a multivariate Gaussian distribution with the diagonal covariance matrix is generally utilized through ignoring the correlation of dimensions due to the low computation and the data storage, the distribution model in traditional speech synthesis is not precise enough. The full covariance matrix must be considered in order to increase the precision of the distribution model. Practically, due to large number of free parameters, the reliable expectation of a full covariance matrix is difficult to estimate. The maximum likelihood linear transformation (MLLT) is adopted to estimate the full covariance matrices to solve the problem.

2.3 EMR Algorithm for Speech Recognition

In research regarding speech recognition, (Tim Barry, 1994) states in his research of the simultaneous use of three machine speech recognition systems to increase recognition accuracy that researchers have been exploring the possibilities of using speech recognition technology to augment the pilot's ability to control and display information in the cockpit. Ensuring a high degree of accuracy in its recognition of pilot commands under all the various noise, vibration and g-force the pilot has to endure are under the hurdles preventing the transition of speech technology into the cockpit.

Furthermore, combining multiple systems into a simple voting architecture may result in an enhancement in overall speech recognition performance. This research use six male and six female from air force military as a volunteer to participate in the experiment to test the vocabulary consisted of the ten digit words (zero to nine) and 10 additional words likely to be used in a simple cockpit digit entry task. Then, he use enhanced majority rules algorithm logic (EMR) software program to determined its own best guess as to which word the subject spoke from the processing the raw recognition and other data received from the three speech recognition devices whenever a word is spoken by the subject. There are shows in the figure 2.3 below the logic of the EMR algorithm when processing the raw data.

In addition, as the procedure going deeper, the subject then "trained" the three speech systems to recognize his or her voice to allowed each recognition

system to collect and store speech templates for use in the recognition portion of the experiment. The EMR systems then read the raw file, computing its own recognition result and adding the recognition result to the raw data file.

From the final result that have been test, he say that from the three speech recognition systems in the EMR algorithm resulted in a better mean adjusted overall accuracy (AOA) than the accuracy obtained by the two individual older-generation systems.

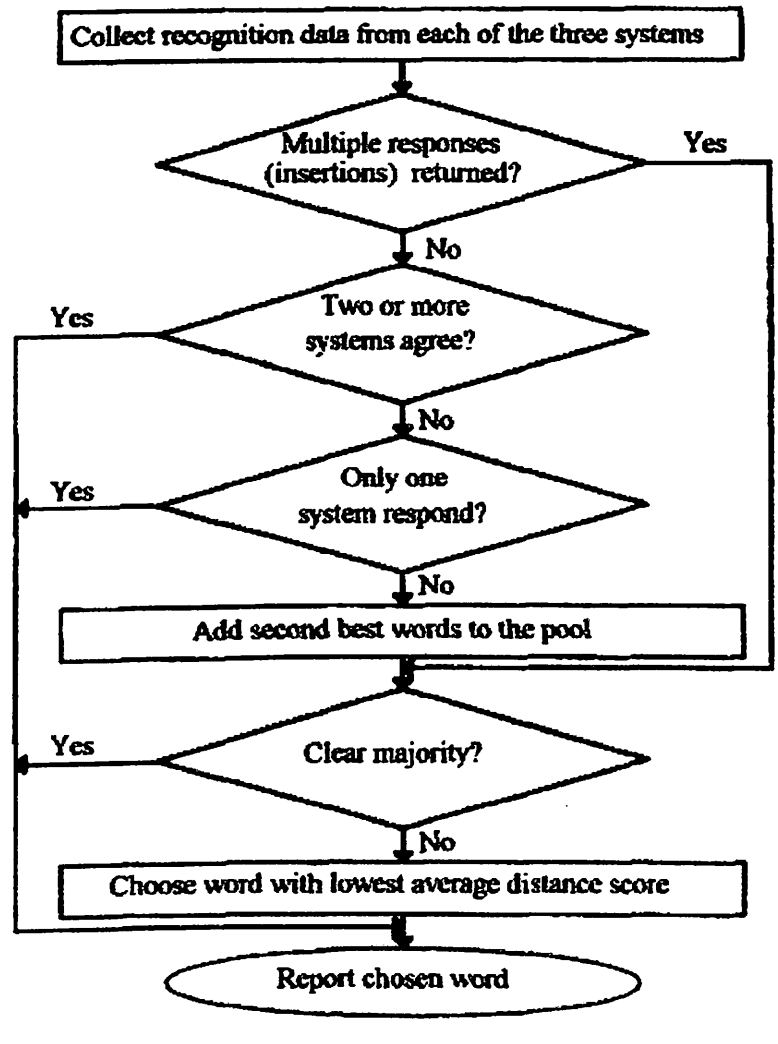

Figure 2.3: Logic flow of EMR algorithm

2.4 Dynamic Approach Training for Speech Recognition

In order to modelling speech synthesis system, several methods had been used in previous research. (Ben Mosbah, 2006). He has done analysis on speech recognition for disabilities people and purposes the suitable features to develop a systems of isolated words recognition and continuous speech recognition systems. While the system of continuous speech recognition systems have like principal application the vocal dictation, the system isolated words recognition have as application the vocal order.

For the disabilities people the absence of the data bases and diversity of the articulatory handicaps are major obstacles for the construction of reliable speech recognition systems, which explains poverty of the market in systems of speech recognition for disabilities people. However, he used two approaches for the continuous recognition systems which are a dynamic adaptation of the phonetic models of the continuous recognition systems for the disabilities people and the used of independent language segmentation (ALISP) for the recognition.

However, training inter speaker variability requires a broad population of speakers in speaker-independent speech recognition systems because it is more important when the speakers have articulatory handicaps. The research concept is to use the phone models from the independent speakers trained on the BREF database to adapt the models each time the system recognizes the sentence correctly.

In conclusion, the research has been done requires continuous evaluation to obtain different results depending on patient's condition. However, this research is expected to help more people with disabilities so that their lives will be easier.

2.5 Automatic Speech Recognition (ASR)

According to (C.Y.Fook, 2012) in his research about Malay speech recognition and audio visual speech recognition, automatic speech recognition (ASR) is an area of research which deals with the recognition of speech by machine in