UNIVERSTI TEKNIKAL MALAYSIA MELAKA
FAKULTI KEJURUTERAAN ELEKTRONIK DAN KEJURUTERAAN KOMPUTER

BORANG PENGESAHAN STATUS LAPORAN
PROJEK SARJANA MUDA II

Tajuk Projek : APPLYING SPEECH RECOGNITION SYSTEM IN LEARNING BASIC ARABIC LANGUAGE

Sesi Pengajian : 2009/2010

Saya **JITVINDER DEV SINGH A/L HARDEV SINGH**
mengaku membenarkan Laporan Projek Sarjana Muda ini disimpan di Perpustakaan dengan syarat-syarat kegunaan seperti berikut:

1. Laporan adalah hakmilik Universiti Teknikal Malaysia Melaka.
2. Perpustakaan dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan dibenarkan membuat salinan laporan ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. Sila tandakan (√) :

☐ **SULIT*** (Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)

☐ **TERHAD*** (Mengandungi maklumat terhad yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan)

☑ **TIDAK TERHAD**

Disahkan oleh:

b/p:

(TANDATANGAN PENULIS)

(COP DAN TANDATANGAN PENYELIA)

LIZAWATI BINTI SALAHUDDIN
Pensyarah
Fakulti Kejuruteraan Elektronik Dan Kejuruteraan Komputer
Universiti Teknikal Malaysia Melaka (UTeM)
Karung Berkunci No 1752
Pejabat Pos Durian Tunggal
76109 Durian Tunggal, Melaka.

Alamat Tetap:

72,PERSIARAN WIRA JAYA TIMUR 50, TAMAN PANGLIMA, 31350 IPOH,PERAK

Tarikh: 30/04/10

Tarikh: 30/04/2010

APPLYING SPEECH RECOGNITION SYSTEM IN LEARNING BASIC ARABIC LANGUAGE

JITVINDER DEV SINGH A/L HARDEV SINGH

This report is submitted in partial fulfilment of the requirements for the award of Bachelor of Electronic Engineering (Computer Engineering) With Honours

Faculty of Electronic and Computer Engineering
Universiti Teknikal Malaysia Melaka

April 2010

"I hereby declare that this report is the result of my own work except for quotes as cited in the references"

Signature : ....

Author : Jitvinder Dev Singh A/L Hardev Singh

Date : ...30/04/10...

"I hereby declare that I have read this report and in my opinion this report is sufficient in terms of the scope and quality for the award of Bachelor of Electronic Engineering (Computer Engineering) With Honours."

Signature : b/p: ....

Supervisor's Name : Puan Afifah Maheran Binti Abdul Hamid

Date : 30/04/2010..............................

Dedicated to my family, specially to my beloved mother, father and sisters, my lectures and lastly my friends

ACKNOWLEDGEMENT

This final year project has given me a lot of experienced gained throughout the process of completing it. I was having many difficulties to complete this project but I managed to pull it through with all the courage and hard work. This could not be done without certain people who helped me in completing this project.

Firstly, I would like to thank **Puan Afifah Maheran Binti Abdul Hamid** who is supervisor whom has given me much guidance to complete this project. She has inspired me to think out of the box and give me ways to look something in a different way. She also gave me a lot of guidance, knowledge and also moral support. My gratitude also goes to my coursemates who also gave a lot of ideas which able me to complete this project.

I also would like to give my special thanks to my parents who gave me support in financial and also moral support. They have helped me a lot and I would not achieve a great success without them. Finally, I would like to thank GOD for guiding me in good and bad times. GOD has shown me the way when I was in the dark moments.

Once again, I would like to give a million thanks to all.

ABSTRACT

Even though there are many learning equipments in the market to learn different languages, there is still lack of correct pronunciation of the language learned. Advances in speech technology and computing power have created a surge of interest in the practical application of speech recognition in learning different languages. By applying speech recognition system, the basic Arabic language pronunciation can be learned in a more effective way. Speech recognition Application Programming Interface (SAPI) will be used to enable the speech processing to compare between the user's voice and pre-recorded voice. Results obtained will enable user to revise their Arabic pronunciations. If the accuracy is equal or greater than 80%, the system will assume that the pronunciation is correct. If the accuracy is less than 80%, the user will be asked to pronounce the word again till the user gets it right. This is a user-friendly system in which it will give feedback to user when user's pronunciation is incorrect. This application is targeted to children in age between 6 to 10 years old. The speech recognition application is developed using Visual Basic.Net software, Praat and Microsoft Speech SAPI software. The main objective of this project is to increase the ability of user's to pronounce the correct basic of Arabic language. By completing this project and making it successful software, more and more children will be able to pronounce the correct Arabic language.

ABSTRAK

Terdapat pelbagai kemudahan belajar bahasa lain di pasaran sekaranag, tetapi masih terdapat kekurangan dalam pembelajaran dalam segi sebutan bahasa yang dipelajari dengan betul. Teknologi yang semakin berkembang dalam "speech recognition" dan juga kuasa komputer telah mengembangkan minat dalam penghasilan alat pembelajaran menggunakan teknologi ini dalam mempelajari pelbagai bahasa lain. Dengan menggunakan "speech recognition", perkataan-perkataan Arab yang asas dapat dipelajari dengan sebutan yang betul. "Speech recognition Application Programming Interface (SAPI)" akan digunakan untuk membolehkan pemprosesan suara untuk membandingkan suara pengguna dengan suara yang telah disimpan dalam sistem. Keputusan yang diperolehi membolehkan pengguna untuk belajar cara bertutur perkataan Arab dengan betul. Jikalau suara pengguna adalah 80% lebih atau sama dengan suara yang disimpan dalam sistem, sistem akan menyatakan bahawa sebutan tersebut adalah betul. Jikalau suara pengguna adalah 80% kurang dengan suara yang disimpan dalam sistem, sistem akan menyatakan bahawa sebutan tersebut adalah tidak betul dan pengguna perlu menyebut perkataan berulang kali sehingga betul. Sistem ini menyasarkan kepada golongan kanak-kanak dalam lingkungan umur 6 hingga 10 tahun. Sistem ini dibuat menggunakan Microsoft Visual Studio, Praat dan Microsoft Speech SAPI. Objektif utama projek ini adalah untuk meningkatkan penguasaan pengguna dalam menyebut perkataan Arab dengan betul. Dengan menyelesaikan projek ini, lebih ramai kanak-kanak dapat mempelajari menyebut perkataan Arab dengan betul.

TABLE OF CONTENTS

LIST OF FIGURE

LIST OF TABLE

LIST OF ABBREVIATION

SAPI	Speech Application Programming Interface
HMM	Hidden Markov Model
DTW	Dynamic Time Warping
GUI	Graphical User Interface
IDE	Integrated Development Environment

APPENDIX LIST

NO	**TITLE**
Appendix A	Main Program
	Category Menu
	Smaller Category Menu
	Learning Category Menu

CHAPTER I

PROJECT OVERVIEW

This chapter will explain briefly about the project background, objectives to be achieved, problem statement and scope of work.

1.1 Introduction

This project is intended to develop a windows application with speech recognition technology to learn basic Arabic. Speech recognition Application Programming Interface (SAPI) will be used to enable the speech processing to compare between the user's voice and pre-recorded voice. The user voice will be received through the microphone. Praat will be used to compare between the user voice signal received and the pre-recorded voice signal in the system. Results obtained will enable user to revise their Arabic pronunciations. A Graphical User Interface (GUI) will be developed as a user interface to allow users to use the software more easily.

Speech recognition is different than voice recognition. In speech recognition, many users can use the software where the software is able to detect different sound signals generated by the different users. Speech recognition is a broad term which means it can recognize almost anybody's speech such as a call centre system designed to recognize many voices. On the other hand, in voice recognition, the software can only be able to detect sound signals generated a user. This means that it can only detect a particular user which is already in the system. Voice recognition is a system trained to a particular user, where it recognizes their speech based on their unique vocal sound.

By using the concept of speech recognition, users which are mainly 6 to 10 years old will be able to learn the correct pronunciation of the Arabic language. They will able to repeat the same process of pronunciation until they are able to say it correctly. Any users can use it because the system is able to process the sound generated by the user. The generated sound will be processed in the system by comparing the pre-recorded sound with the user generated sound. It will repeat the process until the required results is achieved.

To make the learning process becomes more interesting and easy, the system will be developed by using an interesting graphical user interface (GUI) that is suitable with children's need. The system will provide users with the illustration and graphical image for each of the word or pronunciation in the Arabic language. From here, the user will get to understand easily and have an idea of what the system is teaching about. Besides that, the system will also provide the user with bilingual notes. This means that the system will use two languages, which are Malay language and Arabic language.

1.2 Project Objectives

In order for the project to success and to be implemented, the following objectives have to be achieved:-

i. To provide an easy way for children in learning the correct pronunciation of Arabic language without needing any help from other people. By using this system, they will be able to learn from their home themselves.

ii. To attract children by using interesting Graphical User Interface (GUI) so that they will be interested to learn Arabic language.

iii. To apply speech recognition technique in the software and to apply a new technology which is Praat in learning Arabic language.

iv. To reduce the cost as a person does not require extra expenditure for paying Arabic class fees to learn the Arabic language.

v. To develop a software that will be able to help user to have a good pronunciation and vocabulary of the Arabic language.

1.3 Problem Statement

In this modern age, many languages such as Arabic language are being forgotten due to the lack of teaching of those languages in the classroom. Learning a language should be done in the early age. Even thought there are many ways to learn the language, there are still less children learning it because of certain problem. One of it is that most software at market nowadays only provides one-way communication. Besides that, there are some difficulties in finding Arabic language class. Moreover, some parents do not have the required money to send their children to the classes. By developing a Speech Recognition System in Learning Arabic Language, it will be easier to master the language in a short period of time. This system will have pre-recorded speech that will be compared with the user speech. This will help the user to have the correct pronunciation and speak the language in the correct way. This system can also be used by the older generation who is interested in the language.

1.4 Project Scope

The scope of this project is to develop speech recognition software to learn basic Arabic language. The software that will be used in this project is Microsoft Visual Basic 2008 and Microsoft Windows SAPI software. The program code will be written in C#

programming language. Microsoft Windows SAPI is software that builds specially for the speech recognition application. These two software will be linked together to obtain a complete system. Software called Praat will also be used to check the sound signal which was pre-recorded with sound signal received from the user.

This system is specially developed for the children below than ten years old. This system provide user with the basic Arabic language with three steps of learning:

i. The system will play the original sound (pre-recorded voice), and hear the teacher for pronunciation verification.

ii. Then, the user will repeat according to the original sound (user input their voice according to the words that appear on the screen). At this time, the system will inspect the pronunciation of that user.

iii. The system will detect and make comparison between the user's pronunciation and pre-recorded voice. If the accuracy is equal or more than 80%, the system will assume that the pronunciation is correct. If it is below than 80%, the system will enable the user to revise the Arabic pronunciation. This will be done using the Praat software.

Software which is free software is also used. The name of the software is WavePad. This software is used to cut the sound which will be kept in the database of the software developed.

To use this system, users only need to prepare a microphone and a speaker. This software is quite different from the other software in terms of their application. This software provides the user with two way communication. This means that the system will give feedback to user when user's pronunciation is incorrect. Comparing to other software, users only speak or repeat again the words but there are no feedbacks from the system to make a correction to the user. So, the users could not detect either they speak in correct pronunciation or not.

By using speech recognition, many users can use the software. If voice recognition is used, only one user which has train the system with the user voice can use it. Speech recognition allows many users to use it without any restriction.

1.5 Thesis Overview

This thesis is divided into five chapters to provide the understanding of the whole project.

The first chapter of this thesis will explain briefly about the project background, objectives to be achieved, problem statement and scope of work.

Chapter 2 describes about the literature review involved to gather information of the project in order to complete the whole project. This study is focused especially on software in applying speech recognition technique and Microsoft Visual Basic 2008 as the main software to write the program.

Chapter 3 will explain about the project methodology approach taken and how the project is implemented. Each achievement, problems arose and selection taken during the project implementation is explained in detail for each stage until the finishing line.

Chapter 4 will display the output from the project which includes the simulation design and the graphical user interface. This chapter will also discuss and analyze about the project and operation of the software such as their programming code.

Chapter 5 will be the conclusion and suggestion to the project. The recommendation for the future project is explained in this chapter.

CHAPTER II

LITERATURE REVIEW

This chapter describes about the literature review involved to gather information about the project. This study is focused especially on the software and application related to the project.

2.1 History of Speech Recognition Technology

Speech is the primary means of communication between people. For reasons ranging from technological curiosity about the mechanisms for mechanical realization of human speech capabilities, to the desire to automate simple tasks inherently requiring human-machine interactions, research in automatic speech recognition (and speech synthesis) by machine has attracted a great deal of attention over the past five decades.

Speech recognition has been attempted for almost as long as there have been digital computers. As early as 1952, researchers at Bell Labs had developed an Automatic Digit Recognizer, or "Audrey". Audrey attained an accuracy of 97 to 99 percent if the speaker was male, the speaker pause 350 milliseconds between words with limited vocabulary. The defense research agency ARPA, in 1971 sponsored a research initiative to develop a speech recognizer that could handle at least 1,000 words and understand connected speech which means speech without clear pauses between each word. The recognizer could assume a low background noise environment, and it did not need to work in real time [1].

By 1976, the most successful system which was developed by Carnegie Mellon University was called Harpy. Harpy was slow (a four-second sentence would have taken more than five minutes to process). It also still required speakers to 'train' it by speaking sentences to build up a reference model. Nonetheless, it did recognize a thousand-word vocabulary, and it did support connected speech. It used hidden Markov models and statistical modeling to extract meaning from speech.

In 2001, Microsoft released a speech recognition system that worked with Office XP. It neatly encapsulated how far the technology had come in fifty years, and what the limitations still were.

2.2 Speech Recognition

Speech recognition is a broad term which means it can recognize almost anybody's speech such as a call centre system designed to recognize many voices. Speech recognition is a process of taking the spoken words as an input to a computer program or software. This sounds, words or phrases spoken by humans are converted into electrical signals, and these signals are transformed into coding patterns to which pronunciation has been assigned. An isolated-word speech recognition system requires that the speaker pause briefly between words, whereas a continuous speech recognition system does not. There are some external parameters that can affect speech recognition system performance, including the characteristics of the environmental noise and the type and the placement of the microphone.

Table 2.1: Typical Parameters Used to Characterize the Capability of Speech Recognition Systems

Parameters	Range
Speaking Mode	Isolated words to continuous speech
Speaking Style	Read speech to spontaneous speech
Enrollment	Speaker-dependent to Speaker-independent
Vocabulary	Small (<20 words) to large (> 20,000 words)
Language Model	Finite-state to context-sensitive
Perplexity	Small (< 10) to large (> 100)
SNR	High (> 30 dB) to low (< 10 dB)
Transducer	Voice-cancelling microphone to telephone

2.3 Speech Recognition Algorithms

2.3.1 Hidden Markov model (HMM)

Modern general-purpose speech recognition systems are generally based on Hidden Markov Models. These are statistical models which output a sequence of symbols or quantities. One possible reason why HMMs are used in speech recognition is that a speech signal could be viewed as a piecewise stationary signal or a short-time stationary signal. That is, one could assume in a short-time in the range of 10 milliseconds, speech could be approximated as a stationary process. Speech could thus be thought of as a Markov model for many stochastic processes. HMMs are popular is