APPLLYING VOICE RECOGNITION TECHNOLOGY IN LEARNING BASIC MANDARIN LANGUAGE

NOOR FAZIDAH BT HUSIN

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

Faculty of Electronic and Computer Engineering
Universiti Teknikal Malaysia Melaka

April 2009

UNIVERSTI TEKNIKAL MALAYSIA MELAKA
FAKULTI KEJURUTERAAN ELEKTRONIK DAN KEJURUTERAAN KOMPUTER

BORANG PENGESAHAN STATUS LAPORAN
PROJEK SARJANA MUDA II

Tajuk Projek	:	APPLYING VOICE RECOGNITION TECHNOLOGY IN LEARNING BASIC MANDARIN LANGUAGE
Sesi Pengajian	:	2008/2009

Saya NOOR FAZIDAH BT HUSIN

(HURUF BESAR)

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:

___________________________ (TANDATANGAN PENULIS)

____________________________________ (COP DAN TANDATANGAN PENYELIA)

Alamat Tetap: 441-2, KILOMETER 6,
KAMPUNG DUYONG,
75460 MELAKA

Tarikh: ………………………..

Tarikh: ………………………..

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

Signature :…………………………………

Author : NOOR FAZIDAH BT HUSIN

Date :………………………………...

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 Honors.

Signature :…………………………………………………

Supervisor's Name :……………………………………………...

Date :……………………………………………...

I would like to dedicate this thesis to my family and somebody special, whose encouragement and support with a great help in completing it.

ACKNOWLEDGEMENT

Firstly, I would like to thank to Allah the Al-Mighty for blessing and guiding me to finish my final year project successfully. I would like to thank my project supervisor, Miss Syafeeza bt Ahmad Radzi for the guidance along the way of completing this project. My thanks also go to all of the people who have been involved directly or indirectly with this project, only God can reply all of your graciousness.

ABSTRACT

The purpose of this project is to build a learning system of Mandarin Educational in an easier and effective way. This system is developed to attract users especially children to learn Mandarin language. This application is targeted to children in age between 6 to 10 years old. The voice recognition application is developed using Visual Basic.Net software and Microsoft Speech API software. The Visual Basic.Net software is used to write the program coding and Microsoft Speech API software is used for voice recognition system. In this application, the system wills playback the readings that have been recorded before. Then, the user will follow the reading according to the playback recorded by the system and lastly the system will detect and find the comparison between the system's reading and user's reading. If the user's reading is 80% and above similar to the system's reading, the system will assume that the pronunciation is correct. Otherwise, the system will do the correction by playback the correct pronunciation. Voice recognition is one of the best way to attract user especially children to learn the right Mandarin pronunciation. Besides that, it also can improve a person's communication skills.

ABSTRAK

Projek ini bertujuan untuk membina sebuah sistem pembelajaran Bahasa Mandarin dengan lebih mudah dan berkesan. Sistem ini dibangunkan hasil kajian untuk menarik minat golongan kanak-kanak mempelajari bahasa mandarin dengan lebih mudah. Aplikasi ini disasarkan kepada kanak-kanak berumur di antara 6 hingga 10 tahun. Sebuah aplikasi pengenal suara akan dibangunkan dengan menggunakan perisian Visual Basic.Net dan Perisian suara API (Application Programming Interface) di mana perisian Visual Basic.Net digunakan untuk memasukkan program dan perisian suara API adalah untuk proses pengecaman suara pengguna. Sistem yang di bina merupakan sebuah aplikasi di mana sistem akan memperdengarkan bacaan asal yang telah di rakam. Seterusnya pengguna akan merakam semula bacaan yang telah diperdengarkan daripada sistem tersebut dan sistem akan mengesan dan membuat perbandingan di antara bacaan asal dan bacaan dari pengguna. Sekiranya bacaan pengguna menghampiri bacaan asal sebanyak 80% ke atas, sistem menganggap sebutan adalah betul. Jika sebaliknya sistem akan membetulkan sebutan yang salah. Pengenal suara adalah satu cara yang baik untuk mendorong pengguna terutamanya kanak-kanak untuk belajar sebutan perkataan Bahasa Mandarin dengan betul disamping meningkatkan kemahiran komunikasi.

CONTENTS

CHAPTER **TITLE** **PAGE**

PROJECT TITLE **i**
DECLARATION **ii**
DEDICATION **iv**
ACKNOWLEDGEMENT **v**
ABSTRACT **vi**
ABSTRAK **vii**
CONTENTS **viii**
LIST OF FIGURES **xi**
LIST OF ABBREVIATIONS **xii**

1 **INTRODUCTION**

1.1 Introduction 1
1.2 Project Objective 2
1.3 Problem Statement 2
1.4 Scope of Work 3
1.5 Thesis Outline 4

II LITERATURE REVIEW

2.1 Voice Recognition 5
2.2 Voice Recognition Perform 8
2.3 Voice Recognition Software 9
2.4 Microphone usability 10
2.5 Benefit of Voice Recognition software 11
2.6 Speech Recognition 11
2.7 SAPI: Microsoft Window Speech API 12
2.7.1 Application Programming Interface (API) 12
2.7.2 API versus ABI 13
2.7.3 SAPI 5 13
2.7.4 API for Text-to-Speech 14
2.7.5 API for Speech Recognition 14
2.8 Graphical User Interface (GUI) 15
2.8.1 Introduction of GUI 15
2.8.2 How a Graphical User Interface works 15
2.8.3 User Interface & Interaction design 16
2.9 Microsoft Visual Basic.Net 18
2.9.1 Visual Basic Language Concept 19
2.9.2 Creating Your First VB Application 19
2.9.3 Project Fundamentals 20
2.9.4 Microsoft Command & Control Engine 21
2.10 Sound in Wave 22
2.10.1 Sound Wave Properties and Characteristic 22

III PROJECT METHODOLOGY

3.1 Introduction 23
3.1.1 Analysis 24
3.2 Equipment Needed 25

3.2.1 Software 25
3.3 Flowchart of Project Methodology 27
3.4 Software development 28
3.5 Compile 28
3.6 Debug 29
3.7 Graphical User Interface Design 29
3.7.1 GUI Design Considerations 30
3.7.2 Amount of Information Presented 31
3.7.3 Step of GUI Development 32
3.8 Writing Visual Basic Project 35
3.8.1 Planning 35
3.8.2 Declaration & link the Voice Recognition Engine 35
3.8.3 Coding of Visual Basic Programming 36

IV RESULTS AND ANALYSIS

4.1 Introduction 38
4.2 Software features 38
4.3 Operation of the system 39
4.4 Graphical user interface for the project 39
4.5 Home Interface 40
4.6 Menu Interface 41
4.7 Introduction to Numbers Interface 42
4.8 Introduction to Family Interface 44
4.9 Introduction to Animals Interface 46
4.10 Introduction to Body Path Interface 48
4.11 Introduction to Transportation Interface 50
4.12 Exit Interface 52

V DISCUSSION AND CONCLUSION

5.1 Introduction 53

5.2 Discussion 53

5.3 Conclusion 54

5.4 Suggestion 54

REFERENCES 55

APPENDIX 56

LIST OF FIGURES

FIGURE	TITLE	PAGE
2.1	SAPI 5 System	14
2.2	Modern GUI	16
2.3	Start Page	20
3.1	The overall methodology flow	27
3.2	Compiling Process	28
3.3	Process of GUI Development	30
3.4	GUI of Voice Recognition	36
3.5	GUI for Selection Menu	37
4.1	Home Interface	40
4.2	Menu Interface	41
4.3	The Number Interface	42
4.4	The Learning Process Interface	43
4.5	The Family Interface	44
4.6	The Learning Process Interface	45
4.7	The Animals Interface	46
4.8	The Learning Process Interface	47
4.9	The Body Path Interface	48
4.10	The Learning Process Interface	49
4.11	The Transport Interface	50
4.12	The Learning Process Interface	51
4.13	The Exit Interface	52

LIST OF ABBREVIATIONS

ABI	-	Application Binary Interface
API	-	Application Programming Interface
ADC	-	Analog-Digital Converter
COM	-	Component Object Model
GUI	-	Graphical User Interface
IDE	-	Integrated Development Environment
PSM	-	Project Sarjana Muda
SAPI	-	Speech Application Programming Interface
TTS	-	Text-to-Speech
VB	-	Visual Basic

CHAPTER I

PROJECT OVERVIEW

1.1 Introduction

This project is focusing on applying voice recognition technology in learning basic Mandarin Language. This mandarin learning system is provided with new functions and a creative product combining software innovations with Mandarin teaching.

This system uses the same effective methods as Mandarin traditional teaching, such as reading after demos, pronunciation and imitations. It is also creates a variety of new learning methods with the voice recognition technique, such as role-playing, free practice and listening. It is representative Mandarin learning system with most advanced technology at present.

To meet the requirement of technology, this system consists of basic Mandarin learning which can be done using Visual Basic.Net. In the learning process, we suggest beginners and learners with poor basics to choose different sections to learn, and other learners can have intensive practice for difficulties in Mandarin language in their own regions. To make convenience and enhance the learning effect for learners, both

Mandarin and English notes and illustrations for all the standard Mandarin characters and words are designed.

This system provides the user with two way communication. This means that the system will give feedback to user whenever the 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, users do not know whether they speak in correct pronunciation or not.

1.2 Project Objective

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

- To understand the basic concept in database programming techniques.
- To familiar with programming using Visual Basic.Net environment
- To improve communication skill in Mandarin Language
- To encourage children to learn Mandarin Language
- To allow people in different age to keen in studying the Mandarin Language
- To produce a modern technology application in language teaching
- To enhance the children's capability in learning Mandarin Language.

1.3 Problem Statement

Nowadays, the Mandarin Language class for children is difficult to find. By this project, it will solve the problem whereby parents will use the basic Mandarin Language in teaching and guide their children by themselves without attending to any Mandarin Language class. This is related with the main purpose of this learning language which is the natural and fluent application of the language in daily life. Besides that, the two way communication learning software by using voice recognition is not available in the markets yet.

1.4 Scope of Work

The scope of this project is to develop a voice recognition technology system in learning basic Mandarin Language. This system consists of basic Mandarin learning which is done using Visual Basic.Net.

This project is started with carrying a research on applying voice recognition technology in learning basic Mandarin Language via the book and internet. Besides that, the lecturer's guidance is also important for the project development.

The scope of this project is intended to develop a window application with voice recognition technology to learn basic Mandarin. A GUI will be developed as a user interface to show the Mandarin words. Voice recognition API will be used to enable the speech processing to compare between the user's voice and pre-recorded voice.

This system will provide user with three processes:

i. Firstly, the systems will playback the readings that have been recorded before.
ii. Secondly, user will follow the reading according to the playback recorded by the system.
iii. Finally, the system will detect and find the comparison between the system reading and user reading. If the user's reading is 80% and above correct, the system will assume that the pronunciation is right. Otherwise, the system will do correction by playback the correct pronunciation.

1.5 Thesis Outline

This thesis is divided into 5 chapters to provide the understanding of the whole project.

The first chapter of this thesis will include the background of the project, objective of the project which needs to be achieved, the problem statement and all the necessary scope of work regarding the project.

Chapter 2 describes about the literature review that has been used to gather information to complete the whole project. This study is focused especially on voice recognition system.

Chapter 3 will explain about the project methodology and the approach on how the project is implemented. Each achievement and selection taken for each stage will be explained in detail until the project is succeeded. This chapter will briefly describe on history, materials to be that we used and how to operate it. The project workflow is also included in this chapter.

Chapter 4 will display the outcome from the project which is including the simulation design. This chapter will also discuss and analyze about the project.

Chapter 5 will be the conclusion and suggestion to the project. The recommendation 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 Voice recognition

Voice recognition is an alternative to typing on a keyboard. Put simply, you talk to the computer and your words appear on the screen. The software has been developed to provide a fast method of writing onto a computer and can help people with a variety of disabilities. It is useful for people with physical disabilities who often find typing difficult, painful or impossible. Voice recognition software can also help those with spell ling difficulties, including dyslexic users, because recognized words are always correctly spelled.

Voice recognition is the process of taking the spoken word as an input to a computer program. This process is important to virtual reality because it provides a fairly natural and intuitive way of controlling the simulation while allowing the user's hands to remain free.

Voice recognition is the technology by which sounds, words or phrases spoken by humans are converted into electrical signals, and these signals are transformed into coding patterns to which meaning has been assigned.

While the concept could more generally be called "sound recognition", we focus here on the human voice because we most often and most naturally use our voices to communicate our ideas to others in our immediate surroundings. In the context of a virtual environment, the user would presumably gain the greatest feeling of immersion, or being part of the simulation, if they could use their most common form of communication, the voice. The difficulty in using voice as an input to a computer simulation lies in the fundamental differences between human speech and the more traditional forms of computer input. While computer programs are commonly designed to produce a precise and well-defined response upon receiving the proper (and equally precise) input, the human voice and spoken words are anything but precise. Each human voice is different, and identical words can have different meanings if spoken with different inflections or in different contexts. Several approaches have been tried, with varying degrees of success, to overcome these difficulties.

Voice recognition implies only that the computer can take dictation, not that it understands what is being said. Comprehending human languages falls under a different field of computer science called natural language processing. A number of voice recognition systems are available on the market. The most powerful can recognize thousands of words. However, they generally require an extended training session during which the computer system becomes accustomed to a particular voice and accent. Such systems are said to be speaker dependent.

Many systems also require that the speaker speak slowly and distinctly and separate each word with a short pause. These systems are called discrete speech systems. Recently, great strides have been made in continuous speech systems. Because of their limitations and high cost, voice recognition systems have traditionally been used only in a few specialized situations. For example, such systems are useful in instances when the user is unable to use a keyboard to enter data because his or her hands are occupied or disabled. Instead of typing commands, the user can simply speak into a headset. Increasingly, however, as the cost decreases and performance improves, speech recognition systems are entering the mainstream and are being used as an alternative to keyboards.

Voice recognition is a computer technology that utilizes audio input for entering data rather than a keyboard. Speaking into a microphone, for example, produces the same result as typing words manually with a keyboard. Simply stated, voice recognition software is designed with an internal database of recognizable words or phrases. The program matches the audio signature of speech with corresponding entries in the database.

Voice recognition programs that require the user to "train" the software to recognize their particular stylized patterns of speech are called *speaker dependent* systems. Individuals commonly use these types of programs at home or at the office. Email, memos, letters, data and text can be input by speaking into a microphone. Some voice recognition systems, called *discrete speech* systems, require the user to speak clearly and slowly and to separate words. *Continuous speech* systems are designed to understand a more natural mode of speaking.

Discrete speech voice recognition systems are widely used for customer service routing. The system is *speaker independent*, but understands only a small pool of words or phrases. The caller is given a choice to answer a question, usually with "yes" or "no." After receiving an answer, the system escalates the caller to the next level. If the caller replies with a unique answer, the automated response is usually, "Sorry, I didn't understand you; please try again," with a repeat of the question and available answers. This type of voice recognition is also referred to as *grammar constrained* recognition.

Continuous speech is a more sophisticated form of voice recognition software, wherein the caller can speak naturally to explain a problem or request a service. This program is designed to pick out key words or phrases and make a statistical best-guess as to what the customer wants. Speaking plainly aids voice recognition in identifying the need. This type of system has a far more intensive database than discreet speech systems and is also referred to as *natural language* recognition.

2.2 Voice Recognition Approaches

The most common approaches to voice recognition can be divided into two classes:

- Template matching
- Feature analysis.

Template matching is the simplest technique and has the highest accuracy when used properly, but it also suffers from the most limitations. As with any approach to voice recognition, the first step is for the user to speak a word or phrase into a microphone. The electrical signal from the microphone is digitized by an "analog-to-digital (A/D) converter", and is stored in memory. To determine the "meaning" of this voice input, the computer attempts to match the input with a digitized voice sample, or template that has a known meaning. This technique is a close analogy to the traditional command inputs from a keyboard. The program contains the input template, and attempts to match this template with the actual input using a simple conditional statement.

Since each person's voice is different, the program cannot possibly contain a template for each potential user, so the program must first be "trained" with a new user's voice input before that user's voice can be recognized by the program. During a training session, the program displays a printed word or phrase, and the user speaks that word or phrase several times into a microphone. The program computes a statistical average of the multiple samples of the same word and stores the averaged sample as a template in a program data structure. With this approach to voice recognition, the program has a "vocabulary” that is limited to the words or phrases used in the training session, and its user base is also limited to those users who have trained the program. This type of system is known as "speaker dependent." It can have vocabularies on the order of a few hundred words and short phrases, and recognition accuracy can be about 98 percent.

Most voice recognition systems are discrete word systems, and these are easiest to implement. For this type of system, the speaker must pause between words. This is fine for situations where the user is required to give only one word responses or commands, but is very unnatural for multiple word inputs. In a connected word voice recognition system, the user is allowed to speak in multiple word phrases, but he or she must still be careful to articulate each word and not slur the end of one word into the beginning of the next word.

2.3 Voice Recognition Software

Voice recognition software programs work by analyzing sounds and converting them to text. They also use knowledge of how Mandarin is usually spoken to decide what the speaker most probably said. Once correctly set up, the systems should recognize around 80% of what is said if you speak clearly. Several programs are available that provide voice recognition. These systems work best on Windows 2000 and Windows XP.

A number of voice recognition programs can be used with Windows, including a basic one that is supplied with Microsoft Office XP and 2003. Most specialist voice applications include a software CD, a microphone headset, a manual and a quick reference card.

There are several main voice recognition programs available:

Dragon Naturally Speaking

- This program is distributed by Nuance. Naturally Speaking is recognized as the market leader and is the alternative most frequently recommended by Ability Net

IBM Via Voice

- This is also distributed by Nuance. It offers good accuracy, but is not as easy to use as Naturally Speaking.

Q pointer

- Q pointer provides good command and control facilities, but is not so good for writing tasks as is makes more recognition errors. It operates differently to Naturally Speaking and Via Voice.

2.4 Microphone Usability

Microphones are necessary for voice recognition to function, and the quality of a microphone can make the different between an unusable voice product and one where voice recognition shines. A good microphone will have half the voice-recognition error rate of a poor microphone.

However, microphones are one of the largest impediments to people using voice recognition. In general, the failure points are:

1. Users don't have a microphone.
2. Users don't have the microphone plugged into the computer.
3. The sound card isn't compatible with the microphone.
4. If the microphone gain is so high that clipping happens then accuracy will be terrible. If it's too low then accuracy won't be as good as possible.