APPLYING SPEECH RECOGNITION TECHNOLOGY IN LEARNING BASIC MANDARIN LANGUAGE

CHUAN WEN SHENG

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 2010

UNIVERSTI TEKNIKAL MALAYSIA MELAKA

FAKULTI KEJURUTERAAN ELEKTRONIK DAN KEJURUTERAAN KOMPUTER

BORANG PENGESAHAN STATUS LAPORAN

PROJEK SARJANA MUDA II

Tajuk Projek : Applying Speech Recognition Technology In Learning Basic Mandarin Language

Sesi Pengajian : 2009/2010

Saya CHUAN WEN SHENG

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**	

(TANDATANGAN PENULIS)

Alamat Tetap:

Tarikh: 30/04/2010

Disahkan oleh:

(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.

Tarikh: 30/4/2010

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

Signature :.....

Author : CHUAN WEN SHENG

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

"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 : b/p :

Supervisor's Name : CIK SYAFEEZA BINTI AHMAD RADZI /
AFIFAH MAHERAN BINTI ABDUL HAMID

Co-Supervisor's Name: ENCIK MUHAMMAD REDUAN BIN
ABD.LAH SANI (R2 TECHNOLOGY SDN BHD)

Date : 30/04/2010

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

ACKNOWLEDGEMENT

From this final year project, I have gained a lot of experience throughout the process of completing it. I have faced 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 my project supervisor, Cik Syafeeza bt Ahmad Radzi and Puan Afifah Maheran Binti Abdul Hamid for the guidance along the way of completing 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 course mates which also gave a lot of ideas to complete this project.

I would like to thank my co-supervisor, Encik Muhammad Reduan Bin Abd.Lah Sani who is working at R2 Technology SDN BHD in Kuala Lumpur. He has teached and advised me about the concept of using the speech recognition in learning basic Mandarin language. He also gave me a lot of ideas and knowledges on how to use some of software for analysis and testing such as Speex and Praat software.

I also would like to give my special thanks to my parents who gave me a support in financial and also moral support. They have helped me a lot and I would not achieve a great success without them.

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

ABSTRACT

The purpose of this project is to build a learning system of Mandarin Educational in an easier and effective way with using the speech recognition. 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. Even though there are many learning equipments in the market to learn different languages, there is still lack of correct pronounciation of the language learned. By applying speech recognition system, the basic Mandarin language pronounciation can be learned in a more effective way. The Microsoft Visual Studio software is used to write the program coding and the Speex software is used for compressing audio data into a smaller format designed for speech. The Praat software is used to compare the parameter and pitching of the system between user reading and system reading. 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 accuracy is equal or greater than 80%, the system will assume that the pronounciation 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. Speech recognition is one of the best way to attract user especially children to learn the right Mandarin pronounciation. Besides that, it also can improve a person's coummunication skills.

ABSTRAK

Projek ini bertujuan untuk membina sebuah sistem pembelajaran Bahasa Mandarin dengan lebih mudah dan berkesan. Sistem ini dibangunkan untuk menarik minat pengguna khasnya golongan kanak-kanak untuk 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 Microsoft Visual Studio.Net di mana perisian Visual C Sharp.Net digunakan untuk memasukkan program adalah untuk proses pengecaman suara pengguna. Perisian Speex digunakan untuk memadatkan data audio kepada format yang lebih kecil khas untuk pengecaman suara. Perisian Praat digunakan untuk membanding parameter dan nada suara sistem di antara pengguna baca dan sistem baca. 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 menyamai atau lebih besar daripada bacaan asal sebanyak 80% ke atas, sistem menganggap sebutan adalah betul. Jika sebaliknya sistem akan membetulkan sebutan yang salah. Pengenal suara adalah salah satu cara yang baik untuk mendorong pengguna terutamanya kanak-kanak untuk belajar sebutan perkataan Bahasa Mandarin dengan betul disamping meningkatkan kemahiran komunikasi.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ABI - Application Binary Interface

ADC - Analog-Digital Converter

API - Application Programming Interface

CELP - Code-Excited Linear Prediction

COM - Component Object Model

CSR - Command Success Rate

DTW - Dynamic Time-Warping

DTX - Discontinuous Transmission

EP - Extreme Programming

GA - Genetic Algorithm

GUI - Graphical User Interface

HMM - Hidden Markov Model

IDE - Integrated Development Environment

NN - Neural Network

PCM - Pulse Code Modulation

PSM - Project Sarjana Muda

SAPI - Speech Application Programming Interface

SWER - Single Word Error Rate

TTS - Text-to Speech

VAD - Voice Activity Detection

VB - Visual Basic

VBR - Variable Bitrate Operation

VoIP - Voice over IP

WER - Withword Error Rate

CHAPTER I

PROJECT OVERVIEW

1.1 Introduction

This project is focusing on applying speech 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. Software called Speex will be used to compress the voice signal which will be received from the user through the microphone. Praat software will be used to compare the user voice signal received with the pre-recorded voice signal in the system. Results obtained from system will enable user to revise their Mandarin pronounciations. A Graphical User Interface (GUI) will be developed as a user interface to allow users to use the software more easily.

With proper adaptation, speech technology allows beginning language users to practice spoken language outside the classroom. Praat software using the Microsoft C Sharp programming allows learners to have a simulated conversation with a computer. Practicing with such programs should help users improve fluency and confidence. Futhermore, the software can provide individual feedback on pronounciation, which is something that is often lacking in the language classroom. Algorithms calculate how much a given pronounciation has deviated from a model, and then give a score on phonetic accuracy.

To make the learning process becomes more interesting and easy, the system will be developed by using an interesting GUI with C sharp programming. The system will provide users especially children age from 6 to 10 years old with the illustration and graphical image for each of the word or pronunciation in the Mandarin language. From here, the users will more interested and able to understand easily the Mandarin language through speech recognition.

Other than that this system also provides the users with two way communication. This means that the system will give feedback to user whenever the pronounciation 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 pronounciation or not.

1.2 Project Objective

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

i. To understand the basic concept in database programming techniques to develop an interesting GUI and link to some modern technology software application.
ii. To provide two way communiacation between the user and computer system.
iii. Using the speech recognition technology to learn the Mandarin language.
iv. To allow user in different age improve the pronounciation in Mandarin Language.
v. To create a database sound system using C Sharp.Net programming to compare with the user sound frequency.

1.3 Problem Statement

Nowadays, it is difficult to find a Mandarin Language class for children. From 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. On the other hand, a wrong pronounciation with Mandarin language will cause a misunderstanding the meaning of the words when in communication with each other. Besides that, the two way communication learning software by using speech recognition is not available in the market yet. Moreover, some parents do not have the required money to send their children to the classes. By developing a speech recognition system in learning Mandarin language, it will be easier to master the language in a short period of time.

1.4 Scope of Work

The scope of this project is to develop a speech recognition technology system in learning basic Mandarin language. This system consists of basic Mandarin learning which is done using Microsoft Visual Studio, Speex and Praat software. Those two software, Speex and Praat software that are linked together with C Sharp.Net programming in the Microsoft Visual Studio software. The Speex software is use to compress the signal to a small size and smaller version in wav form. The Praat software will check the compressed signal which was pre-recorded with sound signal received from the user.

This project is intended to develop a window application with speech recognition technology to learn basic Mandarin. A GUI will be developed as a user interface to show the Mandarin words.

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 pronounciation is correct. Otherwise, the system will do correction by playback the correct pronounciation.

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

Chapter 2 describes about the literature review that has been use to gather the information to complete the whole project and involved the definition of the algorithm and some examples of the traditional algorithm and modern algorithm based on the speech recognition.

Chapter 3 will explain about the project methodology 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 GUI. 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

2.1 Speech Recognition

Speech recognition is the ability of a machine or program to identify words and phrases in spoken language and convert them to a machine-reable format. Rudimentary speech recognition software has a limited vocabulary of words and phrases and may only identify these if they are spoken very clearly. Speech recognition applications include call routing, speech-to-text, voice dialing and voice search.

The speech recognition sometimes refer to the 'voice recognition' which is the recognition system for trained to a particular speaker. The terms 'speech recognition and 'voice recognition' are sometimes used interchangeably. However, the two terms mean different things. Speech recognition is used to identify words in spoken language. Voice recognition is a biometric technology used to identify a particular individual's voice. Other than that, the speech recognition is the process of converting an acoustic signal that have been captured by a microphone to a set of words. The recognized words can be the final results, as for applications such as commands and control, data entry and document preparation.

In other words, 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. Recognition is generally more difficult when vocabularies are large or have many similar-sounding words. When speech is produced in a sequence of words, language models or artificial grammars are used to restrict the combination of words.[1]. Table 2.1 below shows that typical parameters used to characterize the capability of speech recognition systems.

Table 2.1: Typical Parameters Used To Characterize The Capability of Speech Recognition Systems.[1]

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

Converting a speech waveform into a sequence of words involves several essential steps. First, a microphone picks up the acoustic signal of the speech to be recognized and converts it into an electrical signal. A modern speech recognition system also requires that the electrical signal be represented digitally by means of an analog-to-digital (A/D) conversion process, so that it can be processed with a digital

computer. This speech signal is then analyzed to produce a representation consisting of salient features of the speech. The speech pattern is then compared to a store of phoneme patterns or models through a dynamic programming process in order to generate a hypothesis of the phonemic unit sequence. Dynamic programming is performed to generate the best match while taking these variations into consideration by compressing or stretching the temporal pattern and by probabilistically conjecturing how a phoneme may have been produced. The latter includes the probability that a phoneme may have been omitted or inserted in the utterance.

The technology of speech recognition often finds applications in speaker recognition tasks as well. Speaker recognition can be classified into two essential modes, speaker identification and speaker verification. The goal of speaker identification is to use a machine to find the identify of a talker, in a known population of talkers, using the speech inputs. Speaker verification aims to authenticate a claimed identity from the voice signal.

The performance of speech recognition systems is usually specified in terms of accuracy and speed. Accuracy may be measured in terms of performance accuracy which is usually rated withword error rate (WER), whereas speed is measured with the real time factor. Other measures of accuracy include Single Word Error Rate (SWER) and Command Success Rate (CSR). Most speech recognition users would tend to agree that dictation machines can achieve very high performance in controlled conditions. There is some confusion, however, over the interchangeability of the terms "speech recognition" and "dictation". Both acoustic modeling and language modeling are important parts of modern statistically-based speech recognition algorithms. Hidden Markov models (HMMs) are widely used in many systems.

2.2 Algorithm

An algorithm generally takes some input, carries out a number of effective steps in a finite amount of time, and produces some output. In other words, an algorithm is also an effective method for solving a problem using a finite sequence of instructions. Originally referred to purely mathematical problems but now used in a wider fields such as the data processing, diagnostic problems and many other fields. Each algorithm is a list of well-defined instructions for completing a task. Starting from an initial state, the instructions describe a computation that proceeds through a well-defined series of successive states, eventually terminating in a final ending state. The transition from one state to the next is not necessarily deterministic; some algorithms are known as Hidden Markov Model algorithms and Dynamic Time-Warping algorithm.[2]

2.3 Definition of Traditional Algorithm

Traditional computer languages, like C, concentrate mainly on the definition of the algorithm and data structure components by using language provided mechanisms to specify type definitions, functions, and algorithm control. The interface is under-defined by header files where function names, parameters, parameter types, and parameter order are specified. This is short of specifying the behavior of the interface. Traditional computer languages are much more suited to defining implementation than they are to defining architecture.

All data are only accessible through their own methods, functions and procedures. This courses the program size to explode and makes the execution speed drop enormous because any data access requires at least one subroutine call instead of perhaps just a single assembler statement.

Traditional programming languages easily define the data structures and the algorithms. There is very limited help in defining the architecture. In fact, there is an assumed architecture, so implicit that most languages don't even define it as a feature. The functions main and plus communicate over a shared address space, memory