VOICE INTERACTIVE GAME USING SPEECH RECOGNITION

AHMAD ZULFADLI BIN AHMAD NAZARI

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

May 2008

UNIVERSTI TEKNIKAL MALAYSIA MELAKA
FAKULTI KEJURUTERAAN ELEKTRONIK DAN KEJURUTERAAN KOMPUTER

BORANG PENGESAHAN STATUS LAPORAN
PROJEK SARJANA MUDA II

Tajuk Projek : VOICE INTERACTIVE GAME USING SPEECH RECOGNITION

Sesi Pengajian : 2007/2008

Saya AHMAD ZULFADLI BIN AHMAD NAZARI

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: 25-01-05,
BANDAR BARU SENTUL,
51000, KUALA LUMPUR

MARDIANA BT BIDIN
Pensyarah
Fakulti Kej Elektronik dan Kej Komputer (FKEKK)
Universiti Teknikal Malaysia Melaka (UTeM),
Karung Berkunci 1200,
Ayer Keroh, 75450 Melaka

Tarikh: 9 Mei 2008

Tarikh: 9 Mei 2008

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

Signature : ....

Author : AHMAD ZULFADLI BIN AHMAD NAZARI

Date : 9 Mei 2008

"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 : MARDIANA BINTI BIDIN

Date : 9 Mei 2008

For my lovely mom and dad

ACKNOWLEDGEMENT

First and foremost I would like to thank to Allah the Al-Mighty for His graciousness I am able to finish my Final Year Project successfully. I would also like to thank my project supervisor, Ms. Mardiana Binti Bidin for helping me along the way of completing this project. I would also like express my gratitude to both of my parents for supporting me and also helping me in my studying. My thanks also go to all of the people who have been involved directly or indirectly with this project, only god can repay all of your graciousness.

ABSTRACT

Speech communication refers to the processes associated with the production and perception of sounds used in spoken language. It has been used since the early age of mankind as the medium to communicate and change information with each other. In this Projek Sarjana Muda (PSM) project, a Voice Interactive Game, which is an application where users are able to interact with the game by using the users' voice, is proposed. English words may be hard to pronounce since there are a lot of pronunciation style and some words may sounds very similar. Wrong pronunciation of English words may cause other people to misunderstand what you are saying and may get you into trouble. Voice interactive game is a great way to encourage users especially the youngsters to learn the correct pronunciation of English words which sometimes being pronounce wrongly all the time. For example, the word 'alter' may be wrongly pronounce as 'outer' or 'otter'. Thus, the Voice Interactive Game is proposed for this project. In this project, user will have to provide inputs that are their voice by pronouncing the words provided. The game is targeted for children aged between 7 to 12 years old. Thus, there are only 15 simple words in the system's library. GoldWave 5.06 and MATLAB 7.0 software is used to capture the user's voice and turn it into digital form. The voice signal will be recognized by a recognizer and checked if the pronunciation is correct or not. The recognizer was programmed by using MATLAB 7.0. The results will be shown in the GUI that was also being built by using MATLAB 7.0 software.

ABSTRAK

Komunikasi suara adalah merujuk kepada proses yang berkaitan dengan penghasilan dan persepsi bunyi yang digunakan di dalam bahasa percakapan. Ianya telah digunakan sejak zaman purba lagi sebagai medium untuk berkomunikasi dan bertukar-tukar maklumat sesama manusia. Untuk Projek Sarjana Muda (PSM) ini, sebuah aplikasi *Voice Interactive Game* akan dibangunkan. Ianya adalah sebuah aplikasi di mana pengguna boleh berinteraksi dengan aplikasi permainan tersebut dengan menggunakan suara pengguna tersebut. Perkataan-perkataan bahasa Inggeris kebanyakannya adalah susah untuk disebut memandangkan terdapat banyaknya gaya sebutan perkataan dan sesetengah perkataan berbunyi seakan-akan sama sebutannya. Sebutan perkataan yang salah boleh menyebabkan orang lain salah faham maksud percakapan anda dan mungkin juga akan membawa kepada masalah. *Voice Interactive Game* adalah satu cara yang baik untuk mendorong pengguna terutamanya kanak-kanak untuk belajar sebutan perkataan bahasa Inggeris dengan betul di samping berhibur. Di dalam projek ini, pengguna akan memberikan input suara mereka dengan menyebut perkataan yang disediakan. Aplikasi permainan ini disasarkan kepada kanak-kanak berumur antara 7 hingga 12 tahun. Oleh itu, hanya 15 perkataan ringkas sahaja yang telah disimpan di dalam sistem ini. Perisian GoldWave 5.06 dan juga MATLAB 7.0 digunakan untuk merekod suara pengguna dan menukarkannya kepada bentuk digital. Sistem pengecaman suara pula telah dibangunkan dengan menggunakan perisian MATLAB 7.0. Paparan aplikasi permainan ini juga telah dibangunkan dengan menggunakan perisian MATLAB 7.0.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ADC	-	Analog-Digital Converter
DTW	-	Dynamic Time Warping
GUI	-	Graphical User Interface
GUIDE	-	GUI Development Environment
HMM	-	Hidden Markov Model
LPC	-	Line Prediction Coding
MFCC	-	Mel Frequency Ceptrum Coefficient

CHAPTER I

INTRODUCTION

1.1 Problems statement

There is still no educational software system that is fully interactive with the users available in the market right now, that is can give examples of the correct pronunciation of the words and can detect whether the pronunciation of the users are correct or incorrect. The development of Voice Interactive Game based from speech recognition system will bring a whole new dimension to the educational software with two way interactions.

Since English language is the mother tongue of the world, it is important for us to learn how to speak in English. While we learn and try to speak, we may wrongly pronounce the English words because some words may have similar pronunciation and do not have recognizable differential. For example, the word ‘alter’ may be wrongly pronounced as ‘outer’ or ‘otter’ since there are similarities between those words. And pronouncing the English words wrongly may get us into troubles especially when what we said offended other people.

Although speech recognition system has been around for about 80 years, it is not being utilized into the educational software systems. Many of the educational software have one way interactions with the user which sometimes can be quite boring and user will lose their interest. So with the development of this Voice Interactive Game, it is hoped that educational software will be an interesting application to use and improves the educational level.

1.2 Objectives

For developing the Voice Interactive Game, there are four objectives listed as below:

1. To create an interactive application that is fun to use and at the same time educational.

2. To expose to the mass about the speech recognition system and its technology.

3. To improve pronunciation in English.

4. To implement Digital Signal Processing as the speech recognizer.

1.3 Scope of work

Since the Voice Interactive Game that is going to be developed is a simple game application, the target user for this application will be targeted for the children between the ages of 7 to 12 years old. The reason I chose this range of children age is that in Malaysia, children from the age of 7 years old will be attending the primary school and finished their study in the primary school at the age of 12. So it is good to start early in their age and train them to pronounce the English words correctly.

The project is still in its research period. According to that, the system library will contain only 15 words as the reference words. And the available words will consist of simple words that are not hard for the children to pronounce it.

This project involves no circuit development at all since this is only a software application. The software involved in this project is GoldWave 5.06 and MATLAB 7.0. The reasons I chose these software because they are easily available, easy to understand and use, have a lot of references either on books or the internet and suitable for speech recognition system.

CHAPTER II

LITERATURE REVIEW

2.1 Speech Recognition

What is speech recognition? Speech recognition is the process of converting a speech signal to a sequence of words, by means of an algorithm implemented as a computer program. Speech recognition applications that have emerged over the last few years include voice dialing, call routing, simple data entry, preparation of structured documents, domotic appliances control and content-based spoken audio search.

The performance of a speech recognition system is usually specified in terms of accuracy and speed. Accuracy is measured with the word error rate, whereas speed is measured with the real time factor.

Most speech recognition users would tend to agree that dictation machines can achieve very high performance in controlled conditions. Part of the confusion mainly comes from the mixed usage of the terms "speech recognition" and "dictation".

Speaker-dependent dictation systems requiring a short period of training can capture continuous speech with a large vocabulary at normal pace with a very high accuracy. Most commercial companies claim that recognition software can achieve between 98% to 99% accuracy (getting one to two words out of one hundred wrong) if operated under optimal conditions. These optimal conditions usually mean the test subjects have:

- matching speaker characteristics with the training data,
- proper speaker adaptation, and
- clean environment (e.g. office space).

This explains why some users, especially those whose speech is heavily accented, might actually perceive the recognition rate to be much lower than the expected 98% to 99%.

Limited vocabulary systems, requiring no training, can recognize a small number of words (for instance, the ten digits) as spoken by most speakers. Such systems are popular for routing incoming phone calls to their destinations in large organizations.

2.2 Speech Recognition Technology

Generally, a speech recognition technology use data processing application in the computer to recognize human's voice. Speech recognition can be categorized to several types:

2.2.1 Isolated speech recognition

Speech recognition system that is consists of only one word being spoken in one recognition process. The speaker speaks the words in order to train and test the system.

2.2.2 Continuous Speech Recognition

Speech recognition system that is consists of continuously and clearly spoken words. The words spoken may be spoken more than one word where the words must be spoken clearly one by one. One of the applications that is using this type of speech recognition is saying the sequence of password for bank account for security measure.

2.2.3 Discreet Speech Recognition

Speech recognition system that is consists of naturally spoken words where there is not much gap between the words.

2.3 Speech recognition handling mode

Speech recognition system has two handling mode, where each of the mode consists of different training system.

2.3.1 Speaker dependent

This system is trained by several speakers and only be able to recognize the speech by the speakers. This speaker dependent handling mode is used in order to achieve the system accuracy in recognizing the words spoken by the speaker.

2.3.2 Speaker independent

This system is able to recognize speech not specifically to the speaker only, but also to other users even though there is no speaker's voice sample in the trained system. At this stage, the system can be used as an application for educational software because of its ability to build a global template to match the user's voice and the trained voice.

2.4 Model of speech recognition system

To understand how speech recognition system works, below is the block diagram of the flow of the speech recognition system.

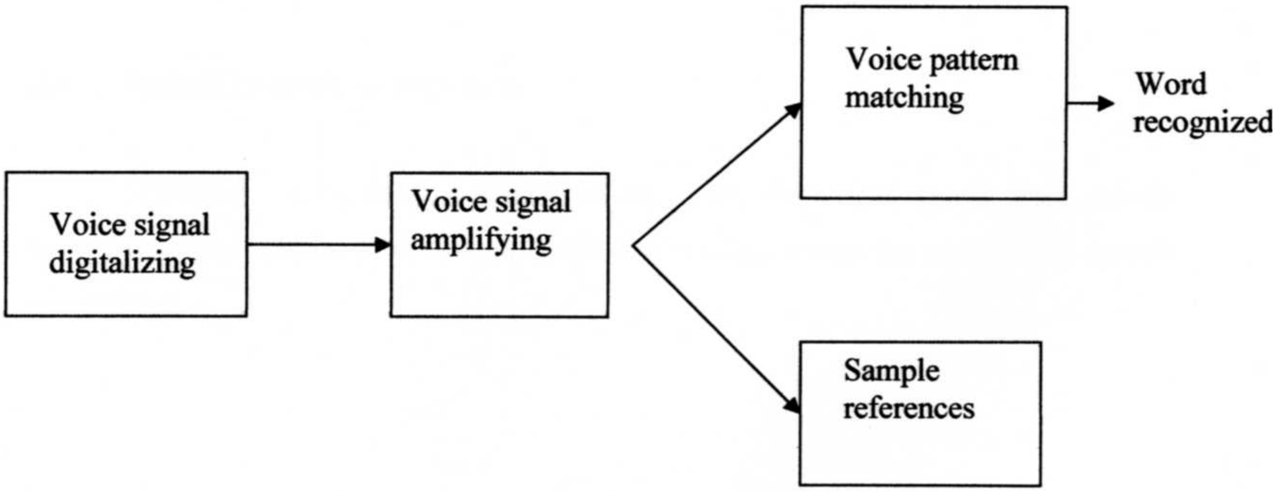

Figure 2.1: Speech recognition system model

Firstly, the input of the user's voice must be converted into digital form. This will allow the voice signal being processed by the computer. It can be done by using a microphone to record the speech and then the speech signal will be converted by

Analog-Digital Converter (ADC) and now we have data representation for every level of discreet time.

After that, the converted voice signal will be processed by using the method of digital signal processing. Form-extracting is a process of taking out as many information that are not related and representing a data that is relevant into a meaningful form. The methods that can be used to form-extract are Fourier transform, Linear Prediction Coding (LPC) and cepstrum. These methods are usually used to produce an array of useful vector form at the pattern-matching section.

Speech recognition system usually needs training process to train the system to recognize certain voice signal. For that purpose, a word will be said over and over again. The voice signal will then be processed and the voice form will be stored as the reference sample.

2.5 Speech recognition approach

Nowadays, there are a lot of products being developed based from speech recognition system. There are two main methods that are use as the approach in speech recognition.