SPEECH RECOGNITION TO TEXT APPLICATION VIA POCKET PC

NURUL AIN BINTI AMINUDDIN

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

Faculty of Electronic and Computer Engineering
Universiti Teknikal Malaysia Melaka

April 2010

SPEECH RECOGNITION TO TEXT APPLICATION VIA POCKET PC

NURUL AIN BINTI AMINUDDIN

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

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 : SPEECH RECOGNITION TO TEXT APPLICATION VIA POCKET PC

Sesi Pengajian : 2009/2010

Saya NURUL AIN BINTI AMINUDDIN
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)

Alamat Tetap:

NO 15, JALAN TAJAK EMPAT, 19/4D,
SEKSYEN 19,
40300,SHAH ALAM, SELANGOR

Tarikh: 19/4/10

(COP DAN TANDATANGAN PENYELIA)
NOOR MAZLINA BT MAHMOD
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: 19 APRIL 2010

"I hereby declare that this report is result of my own effort except for works that have been cited clearly in the references."

Signature : .....

Author : Nurul Ain Binti Aminuddin

Date : 19/4/10

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

Signature : .....

Supervisor's Name : Puan Noor Mazlina Mahmod

Date : 19 APRIL 2010

I dedicate this for my beloved parents

Aminuddin Bin Bee and Fatimah Binti Mamat

My siblings

Mohd Fadzlin Aminuddin and Nurul Izzati Aminuddin

And my entire friend that help me a lot

ACKNOWLEDGEMENT

Thanks to Allah because for completing this "Projek Sarjana Muda". I also appreciate all advice and support from my supervisor, Puan Noor Mazlina Mahmod, those who are very patient in guiding me until I finished. To my parents, thank you very much because both of you always with me through difficulty and good times. Not forget to my friends that gave me idea and helped me from the beginning until the end. May Allah bless my entire friend. Last but not lease, to the Faculty of Electronic and Computer Engineering (FKEKK), because giving me chance to study and complete my project as part of the Bachelor program in Universiti Teknikal Malaysia Melaka (UTEM).

ABSTRACT

Speech recognition has been implementing in computer. There is a lot of PC-based software in market. The difference between each other is either the software is accurate or not. For mobile phones, there are limited brand and type that implement the speech recognition technology for text message service such as blackberry smartphones. However, the speech recognition to text message service is only available between the same smartphones brand and the service is limited and can't be use for other mobile phone's type. Therefore, this project has developed for speech recognition to text message service which can be implemented in any Pocket PC of any brands and types. The main focus is to give a convenient way to type a message through speech recognition technology for people who use pocket PC while they are driving and for people with disabilities to reply text using their fingers or stylus. This project also allow Pocket PC's user to communicate more convenient and efficient without wasting time to type. The project is quite challenging as the speech recognition to text message for pocket PC is still a new issue and a lot of research needs to be done since there are no available software existed yet for speech recognition of Pocket PC-based. The development of this project is believed to bring a great benefit in mobile communication technology nowadays.

ABSTRACT

Speech recognition has been implementing in computer. There is a lot of PC-based software in market. The difference between each other is either the software is accurate or not. For mobile phones, there are limited brand and type that implement the speech recognition technology for text message service such as blackberry smartphones. However, the speech recognition to text message service is only available between the same smartphones brand and the service is limited and can't be use for other mobile phone's type. Therefore, this project has developed for speech recognition to text message service which can be implemented in any Pocket PC of any brands and types. The main focus is to give a convenient way to type a message through speech recognition technology for people who use pocket PC while they are driving and for people with disabilities to reply text using their fingers or stylus. This project also allow Pocket PC's user to communicate more convenient and efficient without wasting time to type. The project is quite challenging as the speech recognition to text message for pocket PC is still a new issue and a lot of research needs to be done since there are no available software existed yet for speech recognition of Pocket PC-based. The development of this project is believed to bring a great benefit in mobile communication technology nowadays.

ABSTRAK

Pengecaman suara telah dilaksanakan pada komputer. Terdapat banyak perisian untuk komputer di pasaran. Perbezaan antara perisian adalah ia nya tepat atau tidak. Untuk telefon bimbit, perlaksanaan pengecaman suara adalah terhad dan hanyalah untuk sesetengah jenis sahaja, contohnya blackberry. Walaubagaimanapun, penggunaan pengecaman suara kepada teks hanya boleh digunakan oleh sesama jenis telefon dan terhad untuk penggunaan telefon jenis lain. Dengan itu, projek ini dilaksanakan untuk pengecaman suara kepada teks yang boleh digunakan pada Pocket PC yang berlainan jenis atau jenama. Fokus utama adalah untuk memberi cara yang lebih mudah untuk menaip mesej menggunakan teknologi pengecaman suara untuk pengguna Pocket PC yang sedang memandu dan orang kurang upaya untuk menghantar mesej menggunakan jari dan stylus. Projek ini juga membenarkan pengguna Pocket PC untuk berkomunikasi dengan lebih mudah tanpa membazir masa untuk menaip. Projek ini agak mencabar sebagai pengecaman suara kepada teks untuk Pocket PC adalah sesuatu yang baru dan kajian yang banyak perlu dilakukan kerana tiada perisian pengecaman suara yang wujud untuk Pocket PC. Pembangunan projek ini adalah berfaedah pada teknologi komunikasi telefon pada zaman ini.

CONTENT

LIST OF TABLE

FIGURE

ABBREVIATION

IVR	-	interactive voice response
ASR	-	automatic speech recognition
DSP	-	digital signal processing
HMM	-	hidden Markov model
MLSA	-	Mel log spectrum approximation
MSD	-	multispace probability distribution
PDA	-	personal digital assistant
USB	-	universal serial bus
VGA	-	video graphics array
GUI	-	graphical user interface

CHAPTER 1

INTRODUCTION

Nowadays, when we call most large companies, a person doesn't usually answer the phone. Instead, an automated voice recording answers and instructs you to press buttons to move through option menus. Many companies have moved beyond requiring you to press buttons, though. Often you can just speak certain words (again, as instructed by a recording) to get what you need. The system that makes this possible is a type of speech recognition program -- an automated phone system. This is the example where they use the speech recognition.

People with disabilities can benefit from speech recognition programs. Speech recognition is especially useful for people who have difficulty using their hands, ranging from mild repetitive stress injuries to involved disabilities that preclude using conventional computer input devices.

First, this project was named voice recognition but after discussed with supervisor, the name changed it to speech recognition because they have a difference. The voice recognition can only recognize one type of voice. In this case, the owner of Pocket PC only can use the software or the application. Meanwhile, the speech recognition can recognize various voices. Then, it can be a user-friendly software because can be use by many user.

The speech recognition for text messages on Pocket PC is beneficial for hands free in a car (i.e. while driving). This software allows the Pocket PC's user sent their text messages using voice. Then the receiver will receive a message in text message. Pocket PC's user didn't have to write or didn't need to touch the screen. They only have to say the message; say the name and the message will send to the receiver.

1.1 Objectives

i. Doing a research about speech recognition.

ii. To design pocket pc application and the application is compatible for all pocket pc.

iii. To implement the speech recognition to text application in Pocket PC and make pocket pc user friendly to driver and people with disabilities.

1.2 Problem Statement

A lot of people nowadays are busy everyday and everywhere they want to use their pocket pc. All people know pocket pc is dangerous while driving. They maybe offend someone is innocence (i.e. student cross road). Focus on pocket pc's user, they hardly to send a text message while driving.

For the people with disabilities, sometimes they hard to communicate using a technology (i.e. people had no hand). They can't type because of these disabilities and they need software that using a voice to communicate.

1.3 Scope

Scope of this project is to design an application that can be use using speech recognition. This application has speech recognition that can recognize any type of voice. Then, after it recognizes the speech/word it will type the message by itself. Then, it can be send to the person that we want to, only by voice.

Then, this application can be use or compatible with any type of pocket pc. This application is to make the pocket pc more user-friendly for people with disabilities. The people that paralytic and blind can use this application. This application is suitable with pocket pc's users who are needs to make a call or sending a text message via their pocket pc at any time while driving by using their voice commands

User	:	A driver and people with disabilities.
Device	:	pocket pc with speech recognition application.
Functionality	:	An application (speech recognition) that can detect a speech and it can write by itself. e.g search detail information from phonebook.
Operation	:	From speech, then convert into text by saying the name or what the user want to say in text using Pocket PC.

1.4 Methodology

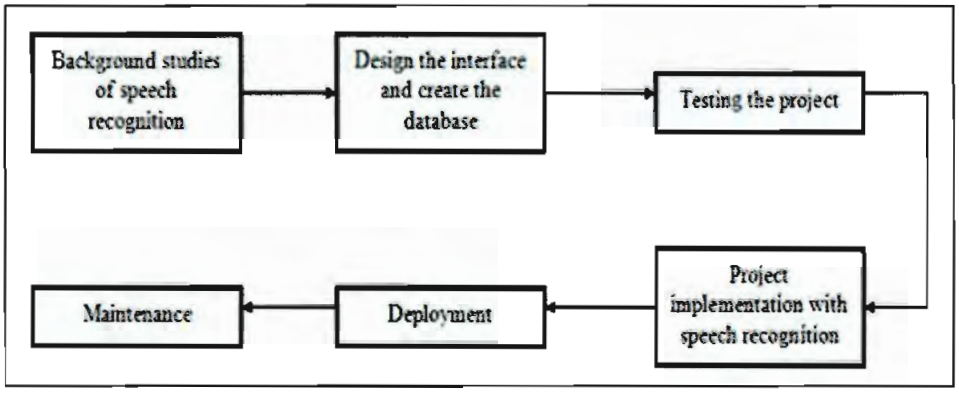

Figure 1: methodology

This project starts by doing a background studies from readability source as books and journals. Then, found certain book that related to the title. The books explain the process of speech recognition and theories.

Designing an interface by using visual basic software, under the smart device then chose pocket PC. Then for the installed templates, chose the device application. After that, design a user interface and create the database using Microsoft SQL server mobile edition.

Check either the database is ok or not. Then, implement speech recognition to Pocket PC. If there have any error, do the maintenance.

CHAPTER 2

BACKGROUND STUDIES

Speech recognition (also known as automatic speech recognition or computer speech recognition) converts spoken words to machine-readable input (for example, to key presses, using the binary code for a string of character codes). The term "voice recognition" is sometimes used to refer to speech recognition where the recognition system is trained to a particular speaker - as is the case for most desktop recognition software; hence there is an aspect of speaker recognition, which attempts to identify the person speaking, to better recognize what is being said.

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. Voice recognition is a system trained to a particular user, where it recognizes their speech based on their unique vocal sound.

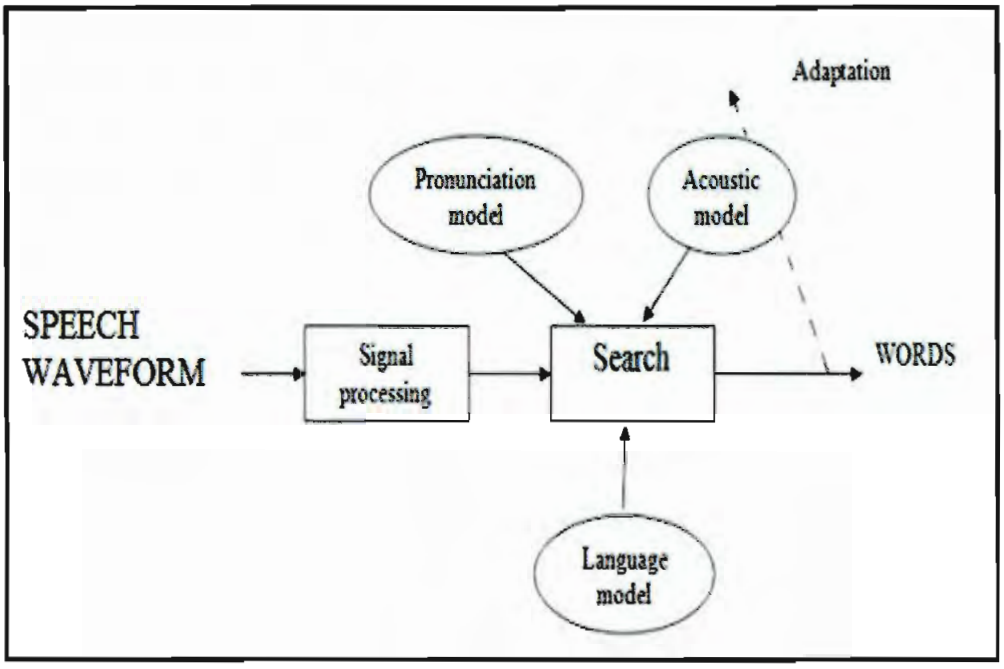

Figure 2: Speech recognition system components

2.1 Speech Signal Analysis

Applications that need voice processing (such as coding, synthesis, and recognition) require specific representations of speech information. For instance, the main requirement for speech recognition is the extraction of voice features, which may distinguish different phonemes of a language. From a statistical point of view, this procedure is equivalent to finding a sufficient statistic to estimate phonemes. Other information not required for this aim, such as the phonatory apparatus dimensions (that is speaker dependent), the speaker moods, sex, age, dialect inflexions and background noise etc.

To decrease vocal message ambiguity, speech is therefore filtered before it arrives at the automatic recognizer. Hence, the filtering procedure can be considered as the first stage of speech analysis. Filtering is performed on discrete time quantized speech signals. Hence, the first procedure consists of an analog to digital signal conversion.

2.1.1 Language model

The language model is file containing the probabilities of sequence of words. A grammar file is a much smaller file containing sets of predefined combinations of words. Language models are used for dictation applications. Whereas grammar is used in desktop command and control or telephony interactive voice response (IVR) type applications.

2.1.2 Acoustic model

Acoustic model is created by taking audio recording of speech and their transcriptions and compiling them into statistical representations of the sound that make up each word.

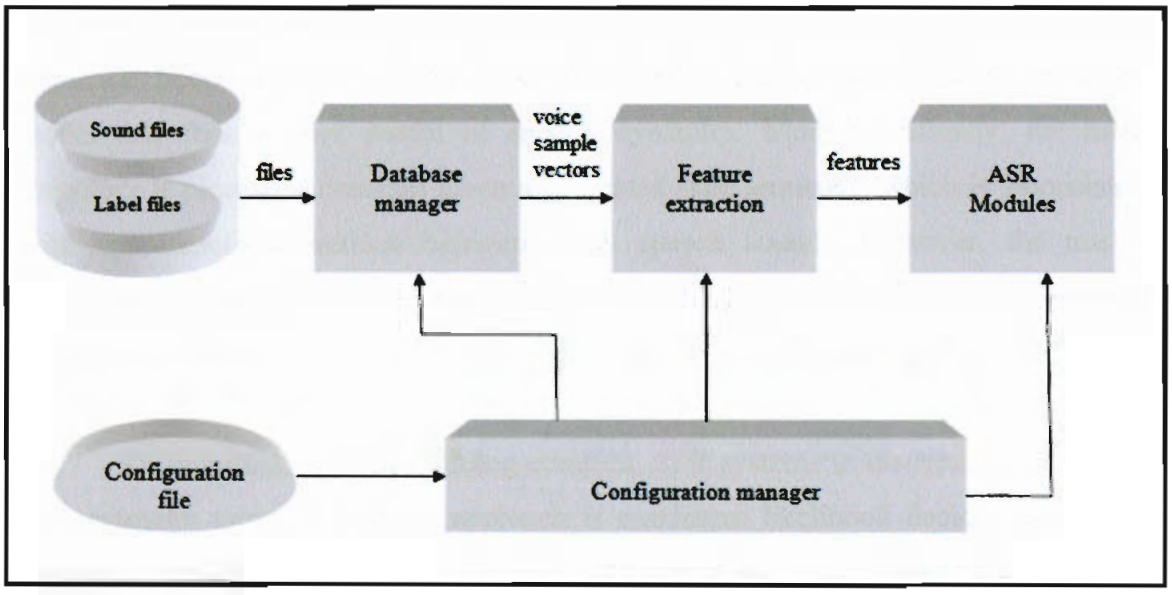

Figure 3: The feature extraction module of an ASR

The procedure aims at mapping data on a more appropriate space where the useless information in discarded. Such a space should depend more on the uttered phonemes than on the actual pitch or speaker or environment. In principle, feature vectors for a given word should hopefully be the same regardless of the way in which the word has been uttered.

The procedure may also reduce the amount of data managed by the ASR. This means that the size of the feature vector may be smaller than of the sample vector. Feature extraction, as well as other digital signal processing (DSP) procedures, can be thought of as a sequence of operations which map input vectors to output vectors. More specifically, vectors of voice samples are the input of a cascade of discrete time systems, here simply referred to as "blocks", to return output vectors according to a specific operation.

The hidden Markov model (HMM) is widely used in speech-recognition systems. The HMM represents acoustics variability in two ways: a hidden, discrete Markov state sequence characterizes temporal variability; and a state-dependent observation distribution characterizes spectral variability, typically using Gaussian mixtures. Observations are assumed to be conditionally independent given the state sequence; hence, HMMs are sometimes described as a piecewise-constant generative model and thus a poor model of speech dynamics. More specifically, the mean trajectory is piecewise constant given a generated state sequence, which is inconsistent with the smooth transitions between many speech sounds. However, the use of derivatives in the observation space (though violating the conditional independence assumption) effectively counteracts this problem, both in recognition and as in synthesis.

An important tool for building complex ASR systems in *distribution clustering* for parameter tying. A popular approach is maximum likelihood decision-tree-based clustering, which is a divisive clustering strategy that uses linguistically motivated candidate splitting questions to provide a mechanism for generalizing to unseen data. Different methods are used depending on the parameter tying assumptions. Clustering can also be used to learn HMM topologies and pronunciation models. Clustering is used extensively in other parts of an ASR system as well, from adaptation to language modeling. It is an important tool for gradually increasing the number of parameters in a model, allowing system complexity to vary as a function of available training data and observed variability.