SPEECH RECOGNITION FOR HOME SECURITY

SARYPAH BINTI HASSAN

This report is submitted in partial fulfillment of requirements for the award of Bachelor of Electronic Engineering (Industrial Electronics) with honours.

Fakulti Kejuruteraan Elektronik dan Kejuruteraan Komputer
Universiti Teknikal Malaysia Melaka

April 2009

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

Signature : …….

Name : Sarypah binti Hassan

Date : …….

"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 (Industrial Electronics) with Honours"

Signature : ……..
Supervisor's Name : Prof Madya Muhammad syahrir bin Johal
Date : ……..

DEDICATION

To my beloved father and mother

Our lecturer and friends

ACKNOWLEDGEMENT

First of all, I would like to take this opportunity to express my deepest gratitude to my beloved project supervisor, Prof Madya Muhammad syahrir bin Johal for his guidance, encouragement and endurance during the whole course of this project. It is indeed my pleasure for his undivided support, invaluable advices and enthusiastic support to make my project a successful one.

Not to forget my beloved family, especially my parents for their fullest support throughout my 3 year's study in Universiti Teknikal Malaysia Melaka (UTeM). It is because of them, I am the person who I am today.

My appreciation to my friends especially my course mates, for their technical advice and material aid. To all the people that assist me directly and indirectly in this project, once again I would like to say a big thank you. Thank you.

ABSTRAK

Pada masa akan datang pengenalan suara merupakan satu teknologi yang dipilih untuk mangawal peralatan serta perkakasan seperti alat mainan, perisian computer dan peralatan electronik. Ia merupakan satu nilai pasaran yang tinggi dan besar apabila teknologi digunakan dengan lebih meluas cume perlu menggunakan arahan suara yang mudah apabila hendak menggunakan peralatan computer,VCR dan sistem keselamatan. System ini mudah dan cepat serta menambahkan kecekapan dengan berkesan. Pengenalan suara merupakan salah satu daripada kaedah biometric. Biometric merupakan pengenalan secara automatik seorang manusia berdasarkan bentuk psikologi dan karakter seseorang. Projek ini mengunakan perisisn MATLAB sepenuhnya serta dibandingkan dengan menggunakan kaedah 'Euclidean distance' dan 'Euclidean squared distance metric'. Sistem ini boleh digunakan dimana-mana sahaja seperti masin ATM, system kedatangan,telefon , kad pintar and perisian computer. Tetapi, untuk projek ini fokus adalah pada peralatan dan penjagaan perkakasan rumah untuk mengelakkan jenayah seperti kecurian dan kecuaian.

ABSTRACT

In the near future speech recognition will become the method of choice for controlling appliances toys, tool, computer and electronic devices. This is huge commercial market just waiting for this technology to mature to control and command an appliance computer, VCR, TV system security system by speaking to it will make easier to use, while increasing the efficiency and effectiveness this project. Speech recognition is one of the leading biometric methods. Biometric is an automatic identification of living person based on physiology or behavioral characteristics. The system of this project to determine a person speaking based on their speech recognition. This project will be implement using the MATLAB software and compare with pattern matching of Euclidean distance and Euclidean squared distance metric. This system is user friendly and could prevent an unauthorized access the fraudulent use of ATMs, Time & Attendance System, Cellular phone, Smart card and computer network. But this project will be focus in home security avoid the criminal is happen. This system can be use at any devices at home to security. This system just a software to show the system is very accurate then the other system.

CONTENT

Chapter 2 LITERATURE REVIEW

LIST OF FIGURE

LIST OF TABLES

LIST OF SHORT FORM

FFT	-	Fast Fourier transform
DCT	-	Discrete Fourier transform
MFFC	-	Mel frequency cepstral coefficient
HMM	-	Hidden Markov Model

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Speech processing is the way of choice for humans to communicate to another person. Speech processing is including a coding, synthesis, recognition, identity verification and enhancement. Based on the human body have a four important part at the human body to appear the vocal tract. The part is larynx for a vocal folds vibrate during voiced sounds, epiglottis to closes off larynx during eating, velum to closes off nose cavity for all sounds except 'm', 'n' and 'ng' and the last part is tongue is used to alter the vocal tract shape. Besides that speech has four sounds category and the category is voiced speech sounds where the vocal vibrate, vowels no blockage of the vocal tract and no turbulence, consonants is a non vowels and plosives the consonant involving an explosion.

Speech recognition is using to convert a speech waveform into text. Speech recognition is one of leading biometric method. Biometric is automatic identification of living person based on physiology or behavioral characteristics. Speech contains information about identity and speaker. A speech signal includes also the language this

spoken, the presence and type of speech pathologies, the physical emotional state of the speaker.

Speech recognition is the system to determine if the person who is speaking is the right person he/she claims to be based on one to one mapping. Speaker recognition can be divided into 2, text dependent and text independent. Recognition system knows text spoken by person in Text dependent since the utterance used in Test and Training set are the same. In text independent, recognition system does not know text spoken by person since the sample speech used in training set is different from test set. Text independent system usually use in forensic cases to detect the criminal or victims.

The system of this project is determines who the person speaking based on their speech recognition with the accuracy. This project is running using the MATLAB software programming and compare with pattern matching Euclidean distance and Euclidean squared distance metric. This system more convenient in relation to the user can potentially prevent unauthorized access to of fraudulent use of ATMs, Time & Attendance System, Cellular phone, Smart card and computer network. But for this project the focus is in home security system.

1.2 Objective

The objective of this project are to implement less computation speaker recognition system with better accuracy, to study the relationship and correlation of the number of training speech with the accuracy and to implement the security system using speech recognition identification. Besides that, speech recognition also to implement less computation speaker recognition system with better accuracy.

1.1 **Problem statement**

The problem with speech recognition is different tones and accents in voices also need to deal intraspeaker variation. The different speaking rate, the emotional state of the speaker and speaking environment are included in the intraspeaker variation. Hence, the sample speeches of the speaker need to be recorded in many different occasions to observe the changes in speaker voice. Besides, the spelling of words doesn't match their sound and the waveform of the word varies a lot between different speaker that mean extract features from speech waveform that are more consistent than the waveform.

Besides that the extracted features won't be exactly repeatable because the characteristics them with the probability distribution. The speech sounds are influenced by adjacent phonemes. Hence, the sample speeches of the speaker need to be recorded in many different occasions to observe the changes in speaker voice. The features of the speeches are then extracted. Choosing the right length of utterance are vital , too short utterance might not contain much information needed for extraction while too long utterance will consume more computation time and sometimes contain information that not needed for extraction.

Somehow, it is needed to experimenting with different length of utterance to get the best identification results. Cepstrum is use to extract the speech feature since it's contain low energy contain in every frames. The lowest coefficients of the Mel Weighted cepstrum are then selected for Pattern Matching process. These lowest coefficients represent the shape of vocal tract of the speaker.

1.3 Scope

This project used the MATLAB software to implement the accuracy of the speech recognition. Then this to convert the signal used the coding to make a feature extraction using the FFT (fast Fourier transform) and convert to the cepstrum signal. This project also used pattern matching to compare the system with better accuracy.

The Pattern Matching process is one of the most important processes in determining a successful identification. Choosing the right Pattern Matching method could affect the success rate of the system. For this project used pattern matching the Euclidean distance metric measure the distance between two feature vectors in decision making phases. The formula of Euclidean, d given by:

$$d(x,y) = \sqrt{\sum_j (x_j - y_j)^2}$$

After that analysis will be performed and get the result from the analysis. The result show the small error differences occur at the last analysis after the comparison. The bar graph will be shown at the last result for each speaker.

CHAPTER 2

LITERATURE REVIEW

2.1 Introductions

This chapter is intent to discuss some fundamental ideas of speaker identification and theory in real word. The features of this project also included here. All the methods using for this project will be explain as well. This chapter also explains about the comparison for this project with other project

2.2 Speech processing

Speech processing is the study of speech signals and the processing methods of these signals. The signals are usually processed in a digital representation whereby speech processing can be seen as the intersection of digital signal processing and natural language processing. Speech processing can be divided in the following categories[8]:

- Speech recognition, which deals with analysis of the linguistic content of a speech signal.
- Speaker recognition, where the aim is to recognize the identity of the speaker.
- Enhancement of speech signals, e.g. audio noise reduction,

- Speech coding, a specialized form of data compression, is important in the telecommunication area.
- Voice analysis for medical purposes, such as analysis of vocal loading and dysfunction of the vocal cords.
- Speech synthesis: the artificial synthesis of speech, which usually means computer generated speech.
- Speech enhancement: enhancing the perceptual quality of speech signal by removing the destructive effects of noise, limited capacity recording equipment, impairments, etc.

Audio signal processing, sometimes referred to as audio processing, is the processing of a representation of auditory signals, or sound. The representation can be digital or analog. The focus in audio signal processing is most typically a mathematical analysis of which parts of the signal are audible. For example, a signal can be modified for different purposes such that the modification is controlled in the auditory domain.

The parts of the signal are heard and which are not, is not decided merely by physiology of the human hearing system, but very much by psychological properties. These properties are analyzed within the field of psychoacoustics.

2.3 Speaker Identification

Speaker identification is a type of speaker recognition. It is the problem of identifying a person solely by their voice. It can be used for purposes such as police investigations. It is different from speaker verification in that, as an example, a criminal's voice is cross checked against a database of criminals' voices looking for a match, and ergo the identity. In contrast, speaker verification seeks to verify, as an example that you really are Mary, seeking to take money out of your bank account using a speaker biometric checking ATM[7].

Speaker identification problems generally fall into two categories:

- Differentiating multiple speakers when a conversation is taking place.
- Identifying an individual's voice based upon previously supplied data regarding that individual's voice.

There are 5 influences to Speaker recognition system accuracy stated in :

I. Phonemes : As the number of phonemes increases, the identification accuracy decreases.
II. Compress Dimension : When the feature vectors are compress into 2 dimensions, a better accuracy could be achieved.
III. Feature Original Vector Dimension : if the original dimension is too high, it is difficult to find vector to differentiate them in low dimension while if its too low the information needed are not enough. So these will affect the accuracy of the system.
IV. Window Function and Interpolation Formula: Commonly window used are Gaussian, square and triangle. Gaussian is selected to guarantee the precision of features' probabilities, and the precision interpolation.
V. The Number of Feature vectors: Too few may affect the probability curve smooth. Too many may not necessary since it will consume much more operation time.

2.4 Speech recognition

In a strict sense, the process of electronically converting a speech waveform (as the acoustic realization of a linguistic expression) into words (as a best-decoded sequence of linguistic units). At times it can be generalized to the process of extracting a linguistic notion from a sequence of sounds, that is, an acoustic event, which may encompass linguistically relevant components, such as words or phrases, as well as

irrelevant components, such as ambient noise, extraneous or partial words in an utterance, and so on[7].

Applications of speech recognition include an automatic typewriter that responds to voice, voice-controlled access to information services (such as news and messages), and automated commercial transactions (for example, price inquiry or merchandise order by telephone), to name a few. Sometimes, the concept of speech recognition may include “speech understanding,” because the use of a speech recognizer often involves understanding the intended message expressed in the spoken words[7].

Currently, such an understanding process can be performed only in an extremely limited sense, often for the purpose of initiating a particular service action among a few choices. For example, a caller's input utterance “I'd like to borrow money to buy a car” to an automatic call-routing system of a bank would connect the caller to the bank's loan department[7].

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 or a microprocessor[7].

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 or a number of hypotheses of the phonemic unit sequence. A phoneme is a basic unit of speech and a phoneme model is a succinct representation of the signal that corresponds to a phoneme, usually embedded in an utterance. A speech signal inherently has substantial variations along many dimensions.

1. First is the speaking rate variation a speaker cannot produce a word of identical duration.