DEVELOPMENT OF SPEECH RECOGNITION SYSTEM FOR FORENSICS APPLICATION

YAP BENG HENG

This Report Is Submitted In Partial Fulfillment of Requirements for the Bachelor Degree of Electronic Engineering (Wireless Communication)

Faculty of Electronics and Computer Engineering
Universiti Teknikal Malaysia Melaka

JUNE 2015

ii

UNIVERSTI TEKNIKAL MALAYSIA MELAKA
FAKULTI KEJURUTERAAN ELEKTRONIK DAN KEJURUTERAAN KOMPUTER

BORANG PENGESAHAN STATUS LAPORAN
PROJEK SARJANA MUDA II

Tajuk Projek : DEVELOPMENT OF SPEECH RECOGNTION SYSTEM FOR FORENSIC APPLICATION

Sesi Pengajian : | 1 | 4 | / | 1 | 5 |

Saya YAP BENG HENG ....(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)

DR. ABDUL MAJID BIN DARSONO
Pensyarah Kanan
Fakulti Kejuruteraan Elektronik Dan Kejuruteraan Komputer
Universiti Teknikal Malaysia Melaka (UTeM)
Hang Tuah Jaya
76100 Durian Tunggal, Melaka

Tarikh: 5 / 6 / 2015

Tarikh: 5 / 6 / 2015

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

Signature :

Author : YAP BENG HENG

Date : 5th JUNE 2015

"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 (Wireless Communication) With Honours"

Signature : Majid

Supervisor's Name : DR. ABDUL MAJID BIN DARSONO

Date : 5th JUNE 2015

To my beloved friends and family

ACKNOWLEGDEMENT

Foremost, I would like to express my appreciation for my beloved family who continuing support me and provide the necessary resources for me to complete my degree final year project.

I also wish to express my sincere thanks to my supervisor DR. Abdul Majid Bin Darsono Head of Department (Telecommunication), for providing supervising and guidance along the way of writing my thesis. Besides, I am feeling thankful to him for sharing his expertise, valuable knowledge in signal processing and encouragement extended to me.

Lastly, I would like to take this opportunity to express gratitude to all my friends who directly or indirectly, have lent their hand for help to develop my project. Thanks to my friend Mr. Chong Tun Hao who assist me by giving idea of coding to develop the system GUI. Thanks to my colleague who always support me mentally and giving encouragement in the process of developing my final year project.

ABSTRACT

Speech recognition system is a system that able to recognize the speech of the speaker. While speaker identification is able to determine the identity of the person speak from the unknown voice. One of the demanding areas of recognition system is in forensics application usually in the case where a crime has been committed and the voice of the criminal needs to be verified from a recorded message. However the sample of voice signal recorded may contain error such as noise which reduces the reliability of this method. The noise occurs in the environment and the status of the person spoke such as illness and drug infection would directly affect the result of the recognition. With the advance development of digital signal processing, more reliable method are found to processed the voice signal which can increase the accuracy of voice matching. In this project, systems which able to determine the identity of the speakers from an unknown voice are develop using MATLAB. A graphical user interface developed so that user is able to interact with the system in a convenient way. The voice feature of the input speech would be extracted by Mel Frequency Cepstral Coefficients (MFCC) compare with the reference model that registers in database. As a result, matching database voice id will be display in the system. From the developed Speech recognition system with GUI, this system should be able to analyze the input voice signal and fairly recognize voice identity of the speaker.

ABSTRAK

Sistem pengenalan suara merupakan sistem yang membantu komputer untuk mengenal pasti perkataan yang diucapkan. Manakala pengenalan suara dapat mengenal pasti suara identiti orang tersebut. Salah satu keperluan dalam sistem pengenalan suara adalah aplikasi forensik. Biasanya kes di mana jenayah yang telah berlaku, mesej suara penjenayah yang telah di rekodkan perlu dikenalpasti . Walau bagaimanapun sampel isyarat suara yang dirakam mungkin mengandungi kesilapan seperti suara sekeliling yang mengurangkan tahap ketepatan kaedah ini. Keadaan yang tidak terkawal berlaku dalam persekitaran dan status orang yang bercakap seperti sakit dan pengaruh dadah secara langsung mengurangkan ketepatan sistem pengenalan suara. Dengan pengunaan pemprosesan isyarat digital, kaedah yang lebih efektif untuk memproses isyarat suara dapat meningkatkan ketepatan padanan suara. Dalam projek ini, satu sistem yang dapat menentukan identiti individu melalui suara dibangunkan dengan menggunakan persisian MATLAB. Ciri-ciri suara itu akan diekstrak mengunakan *Mel Frequency Cepstral Coefficients* (MFCC) untuk membandingkan dengan model rujukan yang ada dalam pangkalan data. Oleh itu, suara pangkalan data mempunyai ciri-ciri yang paling hampir dengan suara rekod akan dikenalpasti identitinya. Melalui GUI sistem pengenalan suara yang dibangunkan, sistem yang direka mampu untuk menganalisa suara dan mengenalpasti identiti suara individu

TABLE OF CONTENT

LIST OF TABLE

LIST OF FIGURE

CHAPTER I

INTRODUCTION

This chapter introduces the project, which included the project background, overview of project, project objective, problem statement as well as the scope of the project.

1.1 Background of the project

Speech is the primary abilities of human being to exchange information with others. Communications in a community are dominated by speech. Because of this, Interactive of speech with the modern technology are require in this current civilization. It is a high demand to implement speech or voice system in the society. Since 60, many research have been done to make the computer able to interpret and understand human language [1]. However speech is a difficult task for digital devices to recognize human voice. Study on technical specifications of sound is requiring in understanding the characteristic of human voice. Through this, digital signal processing method is require filtering and analyzing the voice signal. Speech recognition are applicable in the real world such as multimedia web portal, identity verification and forensic analysis [2]. In forensic analysis, there are several recognition technique used to identify the suspect contribute in criminal cases such as biometric analysis which include fingerprint and voiceprint. By the analysis of a recorded message obtain in a criminal cases, the possible criminal could be identify. As the result it helps legal authority to enforce the law, protect property, and limit civil disorder. Figure 1.1 below show the basic idea for speech recognition, where the unknown voice was compare to the references model. The highest matching score obtained from database compare to unknown voice recognize as the identity of speaker.

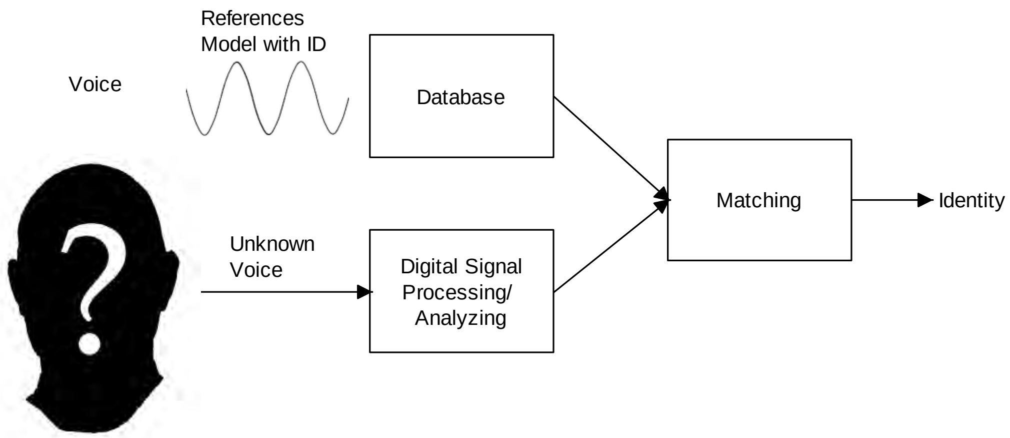

Figure 1.1 Typical Voice Recognition Process

1.2 Problem Statement

In forensic application, voice of the criminal would obtain in a recorded message. These criminal voice needs to be identified and verified However, identification of speaker using speech may not be totally reliable and accurate. Content of recorded message may be blur and not clear .Besides, the Information of speech extracted may affected by the environment (noise) and the speaker status such as ill, influence of drug and so on. In this project, speech recognition with digital signal processing method is use to filter and analyze the content of the recorded voice.

1.3 Project Objective

There are four main objectives in this project which are:

I. To determine the suitable digital signal processing method use in speech recognition system
II. To develop a recognition system that able to determine the identity of the speaker according to the uniqueness of individual voice.
III. To build a graphical user interface for the system by using MATLAB
IV. To analyze the efficiency and reliability of the digital signaling technique use in speaker recognition system.

1.4 Scope of Work

In this project, speech recognition for forensic application is developed by using MATLAB based on digital signal processing technique. To achieve this purposed, references models with ID were collected to comparing the feature of unknown person voice. For this, a database for the system which up to 10 persons is recorded, an ID is assigned for each person. This system should fairly recognize the identity of the speaker based on references model. A graphical user interface for recognition system build using MATLAB to provide interactive task for the user, while the user able to operate the system without the knowledge requirement of the coding.

1.5 Report outline

This report describes the full project development of speech recognition system for forensic application. There are 5 chapters included in this report.

The first chapter of the report introduces flow of the project which includes background of speech recognition, the project objective, problem statement, and scope of the project.

Follow by chapter 2, the relevant studies of previous development of this project. Review of the literature related to the technologies, problem face, technique use further discuss in this chapter. Besides the literature review would contribute to basic idea of the methodology implement based on project objective.

In chapter 3, methodology used to develop this project would be discussed in full detail. This chapter shows the overall flow of the project based on the chosen technique and the practice of the system in real time.

While chapter 4 shows the result of the project. Which include the operation of system, result data tabulate and analysis of the data obtained as well as the recognition system developed.

Lastly, chapter 5 concludes the overall project according the analyses of data obtain with the review of the theory and relevant studies. Comment on the practicality of this system in real word.

CHAPTER II

LITERATURE REVIEW

This chapter briefly describes the concept and theory of speech recognition based on the journal study which includes speech production, biometrics and speech in digital signal processing.

2.1 Speech production

Speech is nature of sound produce by the human voice mechanisms. Voice speeches introduce to the medium through vibration and transmit to listener. The frequent of vibration of these voices in air medium is known as fundamental frequency. The more frequent the vibration of voice represent in higher fundamental frequency or known as pitch, while lower vibration show the lower pitch. Different person would have different speech pitch. Typically, adult female voice average fundamental frequency is higher compare to male.

- Typical adult male speech average pitch from 85 to 180 Hz
- Typical adult female speech average pitch from 165 to 255 Hz

The bandwidth allocated for a single voice-frequency transmission channel is usually 4 kHz for human being. The require sampling frequency for the system at least twice the maximum range of fundamental frequency based on Nyquist–Shannon sampling theorem.

2.1.1 Speech signal classification

Speech recognition system can be separated in different classes by describing what type of utterances they can recognize which include isolated words, connected words and continuous words. Isolated word is referring to single word. While Connected word system are similar to isolated words but allow separate utterance to be run together minimum pause between them. Lastly, Continuous speech recognizers allows user to speak almost naturally, while the computer determine the content. Recognizer with continues speech capabilities are some of the most difficult to create because they utilize special method to determine utterance boundaries

2.2 Biometrics

Biometrics is referring to the technology and technique use analysing biological data. These technologies normally contribute to biological identification and verification. One of the practical applications in this field is forensic application. That suspect and criminal were identified through the scientific procedure. Biometric technique based on intrinsic characteristic, such as fingerprint, finger vein, face pattern and voice [3] .This is because this biological attributes would not lost and forgotten physically over the time. Definitely this method would help during the process of identification as well as verification.

2.2.1 Speech based on biometric

Different person have different voice characteristic. This speciality of voice may refer to the biological and physical condition of individual voice mechanism. Human voice is produce when the air flows from the lung to the throat, vocal and mouth. The sound produced is filter through these voice mechanism. Because of this, combination of these properties consider to a uniqueness of speaker voice [3]. While different shape of the lips may refer to different speech produce. During a biometric investigation, these properties of sound is analyse using the references model pre-recorded to obtain the desired data.

2.3 Speech analysis

The information grabs form the speech signal contributes to identification of speaker [2]. It is the requirement to identify the characteristic of each individual voice which include vocal tract, excitation source and behavior feature for further processing. Figure 2.1 shows the waveform and spectrum of speaker 1. Figure 2.2 shows the spectrum of same sentences spoke by two different speakers. From the spectrum analysis, the peak for the frequency is different from each other, even the same word is mention. In other word, it means that various speaker have peaks at certain frequency [1]. It is the requirement for the system to analyze the characteristic of spectrum distribution. While Figure 2.3 compare the different of voice spectrum of speaker 1 and 2. It shows that, the spectrum amplitude is different among the speaker at a certain frequency.

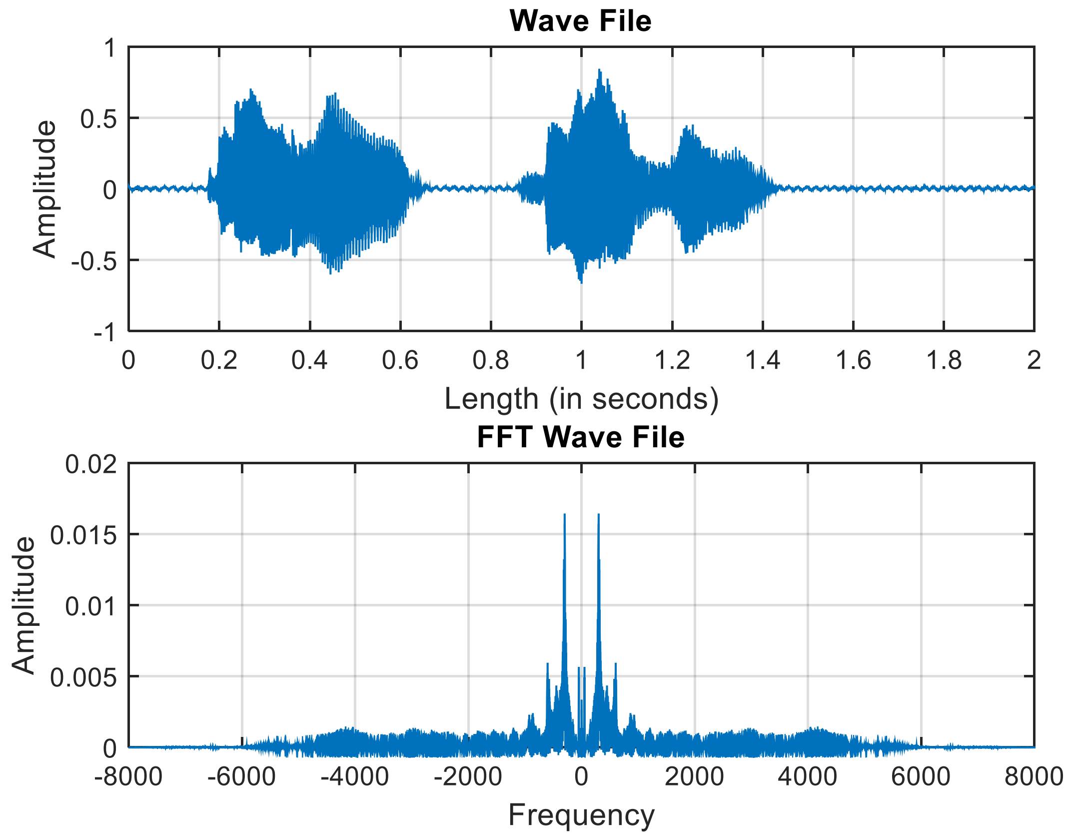

Figure 2.1 Spectrum by speaker 1: Raveena1