

AUTOMATIC SPEAKER RECOGNITION SYSTEM FOR FORENSIC
APPLICATIONS

SHAIFUL ADLI BIN YAAKOB

This Report Is Submitted In Partial Fulfillment Of Requirements For The Bachelor
Degree of Electronic Engineering (Electronic Telecommunication)

Fakulti Kejuruteraan Elektronik dan Kejuruteraan Komputer
Universiti Teknikal Malaysia Melaka

JUNE 2013



UNIVERSITI TEKNIKAL MALAYSIA MELAKA
FAKULTI KEJURUTERAAN ELEKTRONIK DAN KEJURUTERAAN
KOMPUTER

BORANG PENGESAHAN STATUS LAPORAN
PROJEK SARJANA MUDA II

Tajuk Projek : AUTOMATIC SPEAKER RECOGNITION SYSTEM FOR
FORENSIC APPLICATIONS
Sesi Pengajian :

1	2	/	1	3
---	---	---	---	---

Saya SHAIFUL ADLI BIN YAAKOB
(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

(TANDATANGAN PENULIS)

Tarikh: 12/6/2013

Disahkan oleh:

(CAP DAN TANDATANGAN PENYELIA)

Dr. Abdul Majeed Bin Darsone

Penyarah

Fakulti Kejuruteraan Elektronik Dan Kejuruteraan Komputer

Universiti Teknikal Malaysia Melaka (UiTM)

Hang Tuah Jaya

76100 Durian Tunggal, Melaka

Tarikh: 13/06/2013


“I acknowledge that this report is entirely my work except summary and passage I have
quoted the source”

Signature : 

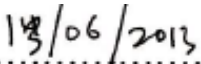
Author Name : SHAIFUL ADLI BIN YAAKOB

Date : 13/6/2013

‘I acknowledge that I have read this thesis and in my opinion this thesis is sufficient in terms of scope and quality for the award of Bachelor of Electronic Engineering (Telecommunication Electronics)’

Signature : 

Supervisor Name : DR ABD MAJID BIN DARSONO

Date : 

I dedicate this report to my father and mother, Yaakob Bin Ismail and Sharifah Zainul Akmar Binti Syed Jaafar, my brothers and my friends for their encouragement and always stand by my side to ensure successfulness

ACKNOWLEDGEMENT

In the name of Allah, the Most Gracious and the Most Merciful. Alhamdulillah, all praises to Allah for the strengths and His blessing in completing this thesis. Special appreciation goes to my supervisor, DrAbd Majid Bin Darsono, for his supervision and constant support. His invaluable help of constructive comments and suggestions throughout the experimental and thesis works have contributed to the success of this project.

I would like to express my appreciation to the Dean, Faculty of Electronics and Computers, associate professor Abdul Rani Bin Othman and also to the panel that has judge me. My acknowledgement also goes to all the technicians and office staffs of Faculty of Electronics and Computers for their co-operations.

Sincere thanks to all my friends especially Miza, Farhan, Faiz, Ismail, Shafiq, Fendi, Fauzan and others for their kindness and moral support during finishing this project. Thanks for the friendship and memories.

Last but not least, my deepest gratitude goes to my beloved parents; Mr. Yaakob Bin Ismail and Sharifah Zainul Akmar Binti Syed Jaafar and also to my brothers for their endless love, prayers and encouragement. For those who indirectly contributed in this research, your kindness means a lot to me. Thank you very much.

ABSTRACT

This report is focus on the application of the automatic speaker recognition system for forensic application and it's called forensic automatic speaker recognition. Forensic recognition aims or applies at the use of individualization. Our voice contains various characterization or parameters that convey information such as emotion, gender, attitude, health and identity. The speaker recognition for this particular project deals with the subject of identifying a person based on their unique voiceprint present in their speech data. There is another important stage that happens before voice feature extraction which called the pre – processing, where it ensures the voice feature extraction contains accurate information that conveys the identity of the speaker. For this particular project, the Mel Frequency Cepstrum Coefficient (MFCC) feature is used to extract the information or the characterization of the speech signal for a text dependent speaker identification system. Vector Quantization- Linde, Buzo and Gray (VQ-LBG) is used to quantized a number of centroids by using this particular algorithm. The codebook of speaker is constituted by these centroids. To be clear, MFCC are calculated in training phase and on the training session. The speaker is identified by using the concept of minimum Euclidean distance of the MFCC of each speaker in training phase to the centroid of individual speaker in the testing phase. All of development of algorithm is performed by using Matlab. The results shows high recognition rate when MFCC is used

ABSTRAK

Laporan ini memberi tumpuan kepada aplikasi sistem pengenalan suara automatik khususnya untuk tujuan forensik. Ia bertujuan untuk mengenal pasti suara individual. Suara manusia terdiri daripada pelbagai ciri perwatakan yang mengandungi informasi seperti emosi, jantina, sikap, kesihatan dan identiti. Pengenal pasti suara untuk projek ini berurusan dengan keadaan di mana ia memerlukan pengenalan pasti seseorang individu berdasarkan ciri-ciri unik yang terdapat dalam data percakapan. Terdapat satu lagi peringkat sebelum peringkat pencerian suara di gunakan. Peringkat ini di namakan peringkat pra-pemproses di mana ia memastikan peringkat pengeluaran ciri suara mengeluarkan maklumat yang boleh membantudalam mengenal pasti pemilik suara yang diuji. Untuk projek ini khususnya, PekaliMel Frekuensi Cepstrum (MFCC) telah digunakan untuk mendapatkan maklumat atau pencerian isyarat ucapan teks bergantung kepada sistem pengenalan. Vektor Pengkuantuman-Linde, Buzo dan Gray (VQ-LBG) digunakan untuk mengkuantumkan beberapa sentroid dengan menggunakan algoritma tertentu. *Codebook* suara di hasilkan oleh sentroid. Selain itu, MFCC dikira dalam fasa latihan dan sesi ujian. Identiti suara dikenal pasti dengan menggunakan konsep jarak minimum Euklidan MFCC setiap penceramah dalam fasa latihan kepada sentroid pengucap suara individu dalam fasa ujian. Semua pembangunan algoritma dilakukan dengan menggunakan Matlab. Keputusan yang diperolehi menunjukkan kadar pengecaman suara yang tinggi apabila MFCC digunakan.

TABLE OF CONTENT

CHAPTER	TITLE	PAGE
	PROJECT TITLE	i
	REPORT VERIFICATION STATUS	ii
	DECLARATION	iii
	DEDICATION	v
	ACKNOWLEDGEMENT	vii
	ABSTRACT	vii
	ABSTRAK	viii
	TABLE OF CONTENT	ix
	LIST OF TABLES	xii
	LIST OF FIGURE	xiii
I	INTRODUCTION	
	1.1 OVERVIEW	1
	1.2 OBJECTIVE	3
	1.3 PROBLEM STATEMENT	3
	1.4 SCOPE OF PROJECT	4
	1.5 PROJECT METHODOLOGY OVERVIEW	5
	1.6 THESIS OVERVIEW	5

II	LITERATURE REVIEW	
2.1	METHODS/TECHNIQUES	7
2.2	FEATURE EXTRACTION	8
2.2.1	Mel-Frequency Cepstrum Coefficient (MFCC)	9
2.3	SPEAKER MODELING	11
2.3.1	Vector Quantization	11
III	METHODOLOGY	
3.1	FEATURE EXTRACTION BY USING MEL FREQUENCY CEPTRUM COEFFIECIENT (MFCC)	14
3.1.1	PRE-PROCESSING TECHNIQUE	15
3.1.1.1	Analogue to Digital Converter	15
3.1.1.2	End Point Detection	16
3.1.1.3	Pre – Emphasis	16
3.1.2	MFCC TECHNIQUE	16
3.1.2.1	Frame Blocking	16
3.1.2.2	Windowing	17
3.1.2.3	Fast Fourier Transform (FFT)	18
3.1.2.4	Mel Frequency Wrapping	18

	3.1.2.5 Cepstrum	19
3.2	FEATURE MATCHING BY USING VECTOR QUANTIZATION LBG	20
3.3	COMPARISON OF TECHNIQUE	23
IV	RESULT & DISCUSSION	
4.1	RESULT	26
	4.1.1 Continuous Speech	26
	4.1.2 Frame Blocking	29
	4.1.3 Windowing	31
	4.1.4 Fast Fourier Transform	33
	4.1.5 Mel Frequency Wrapping	34
	4.1.6 Cepstrum	36
	4.1.7 Result Based On Power Spectrum	42
V	CONCLUSION	
5.1	CONCLUSION	50
5.2	FUTURE WORK	52
	REFERENCES	54

LIST OF TABLES

TABLE	TITLE	PAGE
3.1	Comparison of Automatic Speaker Recognition Technique	23

LIST OF FIGURES

FIGURE	TITLE	PAGE
1.1	The General Concept of Speaker Recognition System	2
2.1	Mel frequency (mels) Against Frequency (Hz)	10
2.2	2 Dimensional of Vector Quantization (VQ)	12
2.3	The Conceptual Diagram of Vector Quantization (VQ)	13
3.1	Block Diagram of Mel-Frequency Cepstrum Coefficient (MFCC)	14
3.2	Block Diagram of Pre-Processing Technique	15
3.3	Example of Windowing to the Speech Signal	17
3.4	Clearly shows linear frequency spacing below 1000 Hz and a logarithmic spacing above 1000 Hz.	19
3.5	Flowchart of Vector Quantization – LBG Algorithm	21
3.6	Steps for LBG Algorithm	22
4.1	Block diagram of Mel Frequency Cepstrum Coefficient (MFCC)	26
4.2	Sample of Speech Signal of long voice sample	27
4.3	Sample of Speech Signal Taken Of Short Voice Sample	28
4.4	Frame blocking of Speech Signal Taken of Short Voice Sample	29

4.5	Frame Blocking of Speech Signal Taken of Long Voice Sample	30
4.6	The Hamming Window	31
4.7	Output Signal of Windowing Process of Speech Signal Taken Of Long Voice Sample	32
4.8	Output Signal of Windowing Process of Speech Signal Taken of Short Voice Sample	32
4.9	Output Signal of FFT Process of Speech Signal Taken of Long Voice Sample	33
4.10	Output Signal of FFT Process of Speech Signal Taken of Short Voice Sample	34
4.11	Filterbank of 20 Filters	35
4.12	Signal after Filter Banks for Long Voice Sample	35
4.13	Signal after Filterbanks for Short Voice Sample	36
4.14	MFCC of Long Voice Sample	37
4.15	MFCC of Short Voice Sample	37
4.16	2D of Acoustic Vectors of Matched Speaker	38
4.17	2D plots of Acoustic Vectors of Mismatched Speaker	39
4.18	Vector Quantization Codebook of Different Speakers (Matched Speaker)	40
4.19	Vector Quantization Codebook of Different Speakers (Mismatched Speaker)	41
4.20	Power Spectrum Plot of the Long Speech Signal	42
4.21	Power Spectrum Plot of the Short Speech Signal	43
4.22	Power spectrum for Different Number of Sample of Long Speech Signal	44
4.23	Power spectrums for Different Number of Sample of Short Speech Signal	45
4.24	Signals Before and After the Mel Cepstrum Filter for Long Speech Signal	46
4.25	Signals Before and After the Mel Cepstrum Filter for Short Speech Signal	47
4.26	Statement of Result Based On Matlab	48

CHAPTER 1

INTRODUCTION

This chapter covers the explanation about the main concept of automatic speaker recognition system for forensic application. Other than that, the main reason why the project is needed is also cover in the objectives of the project. Besides that, what the project covers is also explained the scope of project. The main challenge of the project is also told in the problem statement of the project. Furthermore, how the project is done is also described in the project methodology overview.

1.1 Overview

This project focuses on the application of the use of automatic speaker recognition system for forensic application. The characteristic of the forensic environment at different level such as the development of hierarchical methodology for forensic automatic speaker recognition system by considering the requirement of forensic science are studied.

The main aim is to obtain the identity of unknown person (individual x) from the crime scene or any evidence that has a link to the unknown person that is being investigated. Furthermore, the unknown source can also be attained from scientific analysis of evidence presented in trial.

The concept of automatic speaker recognition can be explained using the Figure 1.1:

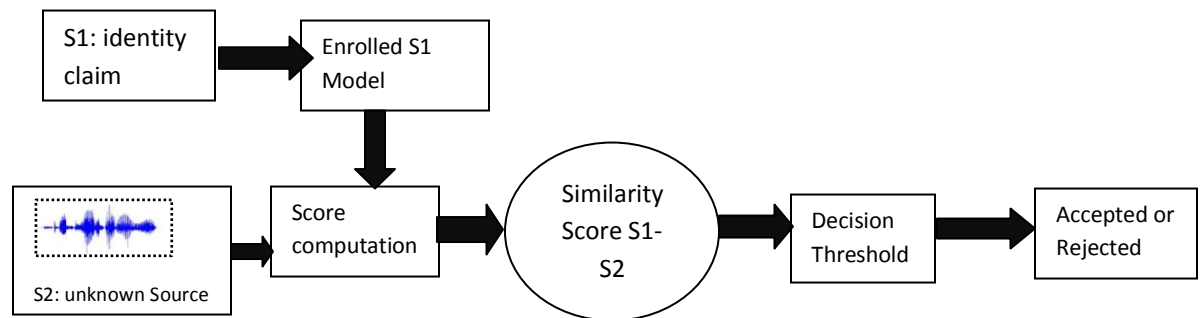


Figure 1.1: The General Concept of Speaker Recognition System

The Figure 1.1 shows that the unknown source is being compared with the known source or someone has their identity claimed as theirs [1-5]. When both of them is being compared, it would result in score computation. This score computation will give the similarity score between the signal that is received from the crime scene and the signal received from the database or the suspect voice.

From the similarity score, the system will decide whether it is accepted (the unknown source and the database source is the same person) or it is rejected (the unknown source and the database is not the same person). If it is to be accepted, the score must be within the decision threshold that has been set earlier on that system.

1.2 Objectives

The objectives of the project are as shown:

- I. To study mathematical of algorithm in development of automatic speaker recognition system. The appropriate algorithm is finding to be used in automatic speaker recognition
- II. To apply the algorithm by using the Matlab Simulink software. Then, we relate the algorithm with in automatic speaker recognition application
- III. To learn an algorithm of automatic speaker recognition system. Thus, we purpose algorithm related to the processing used in audio source separation and apply it through Matlab software.

1.3 Problem statement

Speaker recognition system has its own challenge when it's involves with speaker identification and speaker recognition. This is where the problem starts to arise. This system is based on the premise that a person's speech exhibits characteristic that are unique to the speaker. Each speech signals that is generated through the training and testing session can be greatly different due to the fact of people voice can change through time, health condition, speaking rates and so on [1].

Furthermore, there is also problem that beyond speaker variability that can make the speaker recognition system is more questionable is the acoustical noise and variations in recording environment for example is the speaker uses different telephone handsets [2]. This type of issue could be resolved by using speech feature extraction and feature matching. In feature extraction there is a processor that is called Mel-frequency cepstrum coefficient processor (MFCC) [3].

Speaker recognition system can also be used as a forensic application [4]. In this particular situation, it deals with the issue of how scientist must report to the judge or

jury their conclusion when speaker recognition techniques are used [5]. Other than that, it could also assist in determining specific individuals (suspected speaker) is the source of a questioned voice recording (trace) [6].

The worth of the voice evidence contributed by the speaker recognition technique is determined by forensic expert's role, but in the end it is still up to the jury or judge whether to use the information as an aid in their investigation or not [7].

1.4 Scope of project

The scopes of this project are:

- I. Design the new algorithm of automatic speaker recognition system in forensic application. We can explore the many algorithms in this application and consequently propose new solutions of speaker recognition
- II. Implement an algorithm of speaker recognition system using Matlab software. We examine the relationship between objective performances with algorithm used by speaker recognition.
- III. Develop new methods or techniques of speaker recognition system such as Mel-Frequency Cepstrum Coefficient (MFCC). We can know the suitability, advantages and disadvantages of MFCC techniques to be used in speaker recognition

1.5 Project Methodology Overview

This particular project is developed for speaker recognition for unknown sources with the known sources. The unknown sources might be from the crime scene itself or from any phone conversation that has been tapped to the local, police department. This unknown source can be assumed as forensic evidence. The forensic evidence can be defined as the relationship between such trace, whose source is unknown and some other material, which was generated by a known source or known as suspect. Usually, both of them related to a given crime or offense [2]. Therefore, this project would substantially important to help the jury give their judge and it is done by the use of feature extraction technique such as Mel- Frequency Cepstrum Coefficient (MFCC) and the feature matching technique such as Vector Quantization LBG (VQ LBG).

1.6 Thesis Overview

The objective of this thesis is to provide understanding, propose and implement appropriate algorithm for feature extraction technique such as Mel-Frequency Cepstrum Coefficient (MFCC) and the feature matching technique such as Vector Quantization Linde, Buzo and Gray (VQ-LBG).

Based on that perspective, the chapters are organized and presented as follows:

- Chapter 1: **Introduction**. This chapter is give a comprehensive overview of feature extraction, feature matching and related to it. This includes the objective and problem statement of this thesis, scope of project, and project methodology overview.
- Chapter 2: **Literature Review**. This chapter presents methods for the automatic speaker recognition in forensic applications. It is described about the recognition rate achieve by using these methods.

- Chapter 3: **Methodology**. In this chapter, investigations are carried out to clarify some algorithms. The first part of the chapter described how these technique that used in automatic speaker recognition in forensic applications. Then, why these technique is chosen in speaker recognition
- Chapter 4: **Result & Discussion**. This chapter describes preliminary result that achieve in this thesis is used in automatic speaker recognition.
- Chapter 5: **Conclusion**. Finally, conclusion of this thesis is obtained including the recommendation for future work in automatic speaker recognition

CHAPTER 2

LITERATURE REVIEW

This chapter covers the explanation on the concept of Automatic Speaker recognition (ASR) system. Basically, the concept of the ASR is divided into two main parts which are the feature extraction and feature matching or speaker modelling. Both part will be describe thoroughly in this chapter.

2.1 Methods/Techniques

The identification task is the main aim for speaker recognition and the main objective for this speaker recognition is to recognize the unknown speaker from a set of a known speaker.

The basic modules that are vital in this automatic speaker recognition system are:

I. Frond-end processing:

This is where the sampled speech signal is converted to a set of feature vector. In feature vector, each sampled signal will be characterized based on its properties of speech [15]. These properties of speech can be used to

distinguish or separate different type of speaker. This front end processing is involved directly during the training and testing phase.

II. Speaker modeling:

Feature data is reduced in this phase by modeling the distribution of the feature vectors

III. Speaker database:

Each speaker model is stored in this phase

IV. Decision logic:

The system will make the final decision about the identity of speaker by comparing the unknown feature vector to all models that have been trained and stored in the database [16-19]. The best matching model will be selected from the database [20].

2.2 Feature Extraction

Feature extraction is the process that characterizes the sampled speech signal to make it unique from each other. In other word, feature extraction converts digital speech signal into sets of numerical descriptors called feature vectors that contain key characteristics of the speaker.

Furthermore, feature extraction is the process obtaining different features of voice signal such as amplitude, pitch and the vocal tract [1].It is a task of finding parameter set obtained from the input voice signal. Besides that, extracted features should have some criteria in dealing with the speech signal such as [5]:

- Stable over time
- Should occur frequently and naturally in speech
- Should not be susceptible to mimicry

- Easy to measure extracted speech features
- Shows little fluctuation from one speaking environment to another
- Discriminate between speakers while being tolerant of intra speaker variabilities

2.2.1 Mel Frequency Cepstrum Coefficient (MFCC)

The most popular and prevalent method that often to be used in voice feature extraction is called MelFrequency Cepstrum coefficient. MFCC is based on the human peripheral auditory system [1-5]. The human perception of the frequency content of sounds for speech signals does not follow a linear scale [2-12]. Due to this fact, each tone with an actual frequency, f measured in Hz, a subjective pitch is measured on a scale called the Mel scale.

The Mel frequency scale is a linear frequency spacing 1000 Hz and logarithmic spacing 1kHz. As a reference point, the pitch of a 1 kHz tone, 40 dB the perceptual hearing threshold, is defined as 1000 Mels [1-5].

The difference between the MFC and cepstral analysis is that the MFC maps frequency components using a Mel scale modeled based on the human ear perception of sound instead of a linear scale [1-5]. The short-term power spectrum of a sound using a linear cosine transform of the log power spectrum of a Mel scale is being used in the Mel frequency cepstrum. The formula for this Mel scale is:

$$M = 2595 \log_{10} \left(\frac{f}{700} + 1 \right) \quad (2.1)$$

where f is the actual frequency (Hz).

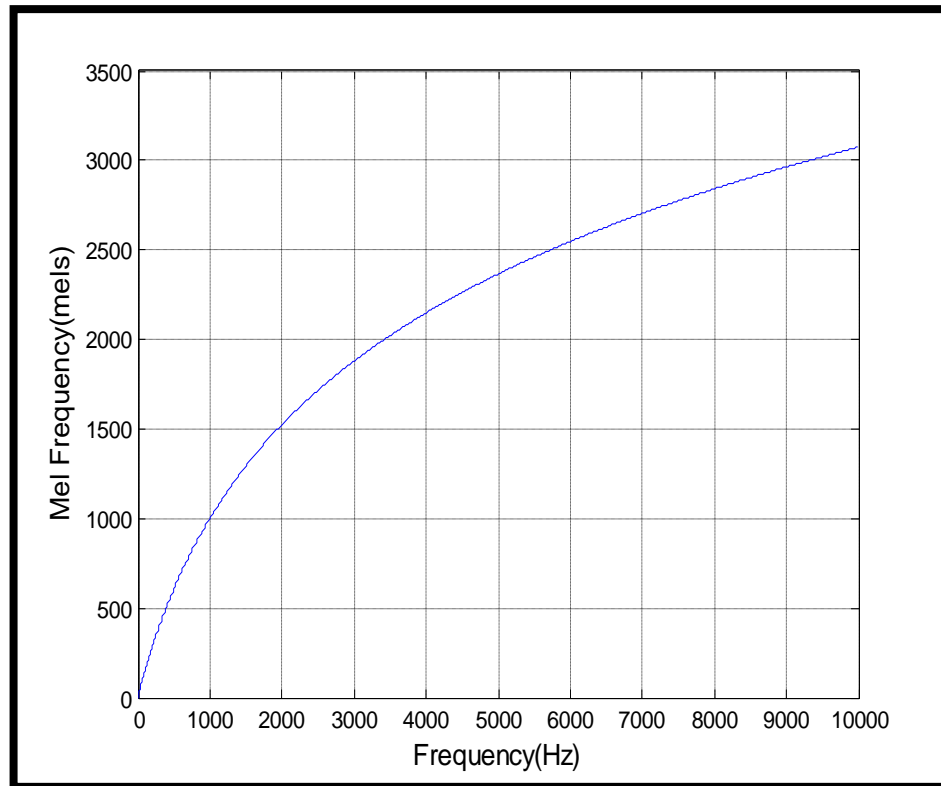


Figure 2.1: Mel Frequency (Mels) Against Frequency (Hz)

From the Figure 2.1, there is obvious evidence that 1000 Hz is equal to the same value of the Mel Frequency (Mels). In another word, if 1000 Hz is the actual frequency, then 1000 Mels will be the result of the conversion between normal frequency value and Mel value. The same concept is applied with the 2000 Hz act as an actual frequency, the result will be 2000 Mels.

The reason that frequency domain parameters are used instead of the normal time domain is due to the fact that the frequency domain parameters are much more consistent and accurate than time domain features. Someone has listed the steps leading to extraction of MFCCs: Fast Fourier Transform, filtering and cosine transform of the log energy vector [5]. The mapping of acoustic frequency to a perceptual frequency scale called Mel scale would produce MFCC.