AUDIO FORENSICS: A VOICE IDENTIFICATION INVESTIGATION

ORYZA SAFUTRA BIN UMAR

A project report submitted in partial fulfillment of the
requirements for the award of the degree of
Master of Information Security

Faculty of Computer Science and Information System
Universiti Technologi of Malaysia

JANUARY 2013

Dedicated to my beloved father, mother, and my little sister
for their prayers, supports, and sacrifices
that make me come to this stage successfully.

To my dearest Nurhayati,
my inspiration who wholeheartedly share all my joy and bitterness,
who provide me his support and faith with no regrets.

ACKNOWLEDGEMENT

"In the Name of Allah, the Most Gracious and the Most Merciful"

Alhamdulillahirabbil „alamin, I would like to deeply praise **ALLAH SWT.** for allowing me passing all of these moments, for giving me the strength and ability to complete this Master Project entitled Audio Forensic: A Voice Identification Investigation.

I wish to express my sincere gratitude to my beloved family, my father, mother, and my sister for their support, kindness and encouragement. They have been so great and I know there would be no way I could have such a good life without having the love and care from them.

My appreciation to Pn. Mazleena Salleh as my supervisor. Thanks for her comments, encouragement, advice and ideas in completing the project. Special gratitude for all of my lecturers, tutors and friends for all these 1,5 years who provide me the assistance either directly or indirectly in completing this project.

ABSTRACT

The development of computer technology has result in demand for more effective intelligent computer program. One of the areas is speaker identification (SI). SI is the process of identifying the speaker based on the characteristics contained in their speech waves. This process can be used in forensic investigation to recognize voice of suspected criminal. Nowadays, a lot of methods can be used to perform speaker identification. Nevertheless, the accuracy of these methods is different according to its algorithm that being used as well as the analyzing of the data. One of methods that can be used for speaker recognition is wavelet transform (WT). WT divided into two methods; discrete wavelet transform (DWT) method and continuous wavelet transform method. This research focused in the implementation, development and analyzing the accuracy of DWT in identifying voice. The experiment is conducted to recognize the spoken person and this is done in four different approaches: recognition based on a single predefined spoken word with normal voice, recognition based on a single predefined spoken word with non-normal voice (with nose closed), recognition based a on multiple spoken words including the predefined word and recognition based on single predefined word but with different tone frequency. The results obtained are voice with changed frequency such as in experiment two and three gives accuracy below 50 percent and for voice with normal frequency like in experiment one and four gives the accuracy above 80 percent. However DWT gives satisfactory result if the voice frequency is normal.

ABSTRAK

Perkembangan teknologi komputer adalah persekitaran yang cepat berubah. Hal ini memerlukan sistem yang lebih ramah pengguna yang salah satu caranya boleh dicapai dengan menggunakan program komputer cerdas. Salah satu bidang adalah penceramah pengenalan (SI). SI adalah proses mengenal pasti penceramah berdasarkan ciri-ciri yang terkandung dalam gelombang pertuturan mereka. Proses ini boleh digunakan dalam penyiasatan forensik untuk mengenali suara penjenayah. Kini, banyak kaedah boleh digunakan untuk melaksanakan pengenalan penceramah. Walau bagaimanapun, ketepatan kaedah ini adalah berbeza mengikut algoritma yang digunakan dan analisis data. Salah satu kaedah yang boleh digunakan untuk pengiktirafan penceramah adalah ubahan wavelet (WT). WT dibahagikan kepada dua kaedah; diskret ubahan wavelet (DWT) dan jelmaan wavelet berterusan. Penyelidikan ini tertumpu dalam pelaksanaan, pembangunan dan menganalisis ketepatan DWT dalam mengenal pasti suara. Eksperimen ini dijalankan untuk mengenal pasti orang yang dituturkan dan ini dilakukan dalam empat pendekatan yang berbeza: pengiktirafan berdasarkan satu perkataan yang dipratentukan dituturkan dengan suara normal, pengiktirafan berdasarkan satu perkataan yang dipratentukan dipertuturkan dengan suara yang tidak normal (dengan hidung tertutup), pengiktirafan berasaskan pada perkataan berganda dituturkan termasuk perkataan yang dipratentukan dan pengiktirafan berdasarkan perkataan yang dipratentukan tunggal tetapi dengan kekerapan nada yang berbeza. Keputusan yang diperolehi adalah suara dengan frekuensi berubah seperti eksperimen dalam dua dan tiga memberikan ketepatan di bawah 50 peratus dan untuk suara dengan kekerapan biasa seperti dalam eksperimen satu dan empat memberikan ketepatan melebihi 80 peratus. Walau bagaimanapun DWT memberikan hasil yang memuaskan jika kekerapan suara adalah normal.

TABLE OF CONTENT

LIST OF TABLE

LIST OF FIGURE

LIST OF ABBREVIATION

ADC	-	Analog to Digital Conversion (ADC)
ANN	-	Artificial neural network (ANN)
ASR	-	Automatic Speech Recognition (ASR)
ATM	-	Automatic Teller Machine (ATM)
BP	-	Back-propagation (BP)
BER	-	Bit Error Ratio (BER)
CWT	-	Continuous Wavelet Transform (CWT)
DAC	-	Digital to Analog Converter (DAC)
DWT	-	Discrete Wavelet Transform (DWT)
FT	-	Fourier Transform (FT)
MFCC	-	Mel Frequency Cepstrum Coefficient (MFCC)
MLP	-	Multi-layer Perceptron (MLP)
PCM	-	Pulse Code Modulation (PCM)
WT	-	Wavelet Transform (WT)
N	-	Frame Size
M	-	Frame Increment
T	-	True
F	-	False

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Overview

Human communication is dominated by speech and hearing. Most quick transfer of information from one person to another is always carried by speech. And that was about the interaction between humans. The man-machine communication, for years, is dominated by typing. However, the man-machine communication can also be done in several ways including writing and speaking along with their disadvantages or weaknesses of the use.

Speech is potentially the fastest form of man-machine communication. Speaking rates vary from about 120 to 250 words per minute (Ainsworth, 1988). This is slightly faster than skilled typing rates. Automatic Speech Recognition (ASR) can be defined as the process of converting an acoustic signal, captured by the microphone, to a set of words understood by the recognition engine. Speech, however, does not need to be learned, or at least it is learned with little effort in early childhood. Speech processing is a rapidly developing field, driven by much expected applications in telecommunication, man-machine interaction, and the like. In fact, speech is a very special type of signal that has received much attention.

Speaker recognition is the process of recognizing the speaker based on the characteristics contained in their speech waves. There are two main fields of speech recognition; speaker identification and speaker verification. This research will focus on the speaker identification. Speaker identification is done by comparing the extracted speech signal from an unknown speaker to a database of known speaker (Price and Eydaghi, 2006). There are two categories of speaker identification; text-dependent and text-independent. In text-dependent speaker identification, the speaker is required to say the same key word or sentence for both training and testing, whereas in text-independent speaker identification does not rely on the specific text to be spoken by the speaker. This research project will be done in text-dependent or text-prompted speaker identification. The system can be easily deceived if the recorded voice of a registered speaker is played, it can be accepted as the registered speaker.

1.2 Problem Background

As speech interaction between human and computers becomes more pervasive in life activities, the utility of automatically identifying a speaker is much developed. Speaking, however, is the fastest way to communicate with machine. Speech is a natural mode of communication for people. It does not have to be learned, but has been a natural ability to learn. Speech leads to a faster time for the solution of a problem. It has many advantages and has been used in many applications that require a fast access or command such as in manufacturing, in aviation, medical application, security, and also a helpful way to communicate for the disabled.

Several factors can cause errors in that voice recognition process through the classification process includes:

1. State of extreme emotional (Stress)
2. Deficiency of the room acoustics (Noise)
3. Different types of recording microphone
4. sickness (such as flu that can change the vocal tract)

From several factors above shows that the state of body or human health may affect the classification results in a voice biometric technology such as speaker recognition.

Nowadays, a lot of methods can be used to perform speaker recognition. The methods that can be used for speaker recognition are:

1. Fourier Transform (FT)
2. Backpropagation (BP)
3. Mel Frequency Cepstrum Coefficient (MFCC)
4. Multi-layer Perceptron (MLP)
5. Wavelet Transform (WT)

BP, MFCC, and MLP are methods that based on Fourier transform, and the identification has reached 100 %. However, Fourier transform has some disadvantage that limit its applicability, there are less able to provide information signals in time domain and frequency simultaneously and analyze the signals are not stationary, so to handle the disadvantages of Fourier transform, other approach in signal processing is needed besides Fourier transform, that is wavelet transform.

Wavelet transform method is a method that begins popular for signal processing, such as images and sound, and wavelet transform has not been widely applied to the analysis of sound, especially for audio-based recognition. Wavelet transform produces a good time resolution at high frequencies in determining the initial parameterization of voice and voice characteristics of short duration and able to analyze the signal discontinuous (non-stationary) accurately.

Wavelet transform is divided into discrete wavelet transform is used when the processed signals are discrete signals and continuous Wavelet Transform is used when the processed signals are continuous signals. This study used discrete wavelet transform because it uses a discrete signal.

1.3 Problem Statement

With the existence of such problems in the 1.2 then there are several important points to perform the solution:

1. Clarity of voice level that will be inputted
2. Process of extracting features of input voice
3. Level of voice recognition that has been extracted.

Based on these three statements, appear the idea to make how to process the three of statement became basic of the solution process. Then the formulations of the problem are:

1. Knowing the process of voice recognition in general advance
2. Analyzing feature extraction method that is the discrete wavelet transform to process the input.
3. Knowing the process of feature extraction methods that have been selected
4. Knowing the accuracy of voice recognition methods that have been selected to perform.

With these fourth formulations of the problem above, the process will be determined in general, how the input will be processed from the stage of receiving input to the stage of getting the desired results, so these fourth formulations of the problem above is a description to know what to do further.

1.4 Aim

The aim of this research project is to define the process and accuracy of Discrete Wavelet Transform algorithm as feature extraction and vector quantitation as feature extraction analyzation in identifying different voice experiment.

1.5 Objectives

The objectives of this research are:

1. Studying and implementing discrete wavelet transform to process the input
2. Analyzing the existing speaker identification method for document
3. Implementing and developing audio-based recognition with discrete wavelet transform method and vector quantization to obtain accurracation of the algorithm.
4. Testing and validating the prototype of discrete wavelet transform method which will be implemented

1.6 Scopes

The scopes of this research are:

1. This research is limited to speech pattern for identification by using discrete wavelet transform.
2. This research is not focused to recognize children speech and/or people with disabilities of speaking.
3. The speech is limited to Malaysia language.

4. The research comparison is limited to some learning parameters (training time and number of epochs) and accuracy of speaker identification.

1.7 Research Justifications

The purpose of this research is to follow up the development of speech processing which is becoming widely used nowadays. By using speech technology in many applications, the work of a particular field will be faster. When speech was involved, many more messages were sent between the participants. By doing this research, hopefully, others can get new knowledge of speech processing in today"s life by using wavelet transform method. This could be useful for people who work in manufacturing or aviation where concurrent tasks and control are needed and for the disabled who needed to be able to control appliances easily by voice. Also, the security system would be enhanced by using ASR since it can replace the written password to a spoken speaker identification one such as in Automatic Teller Machine (ATM) system, security system, and also can be used in forensic investigation to recognize the voice of suspected criminal.

REFERENCES

Agustini Ketut. *Biometrik Suara Dengan Transformasi Wavelet Berbasis Orthogonal Daubenchies*, pp 1-5.

Ainsworth, W. A. (1988). *Speech Recognition by Machine*. UK: Peter Peregrinus Ltd.

Alfatwa Fathony Dean. *Watermarking pada Citra Digital Menggunakan Discrete Wavelet Transform*, pp 4-8.Bandung.

Bose *N*. K., Liang P. (1996) *Neural Network Fundamentals with Graphs, Algorithms, and Applications* (McGraw-Hill, New York).

Campbell, J. (1997). *Speaker Recognition* : A Tutorial.___. IEEE.

Chris Rowden. Speech Processing, *McGraw-Hill International Limited*, 1992.

Darma Putra. (2009). *Sistem Biometrika. Konsep Dasar, Teknik Analisis Citra, dan Tahapan Membangun Aplikasi Sistem Biometrika*. Yogyakarta.

Handoko, Santi K. dan Adikusuma, Tan I. (2005). *Disposal management system PT "x". (TA No.02040972/IND/2005).* Unpublished undergraduate thesis, Universitas Kristen Petra, Surabaya

Honda, *M*. (2003). Human Speech Production Mechanisms. *NTT Technical Review.* 1(2), 24-29.

Krishnan, J. (1994). Auditor swiching and conservatism. *The Accounting Review 69*: pp.200-215.

Li Guohui, A.Khokhar Ashfaq. *Content-based Indexing and Retrieval of Audio Data using Wavelets*, pp 2-3. Newark.

Mahmoud, *M*. I., Dessouky, *M*.I.*M*., Deyab, S. and Elfouly, F.H. (2007) Comparison between Haar and Daubechies Wavelet Transformions on FPGA Technology. *World Academy of Science: Engineering and Technology 26.*

Mallat, S. (1999): *A Wavelet Tour of Signal Processing, Academic Press*, USA.

Manunggal, H.S. (2005). *Perancangan dan Pembuatan Perangkat Lunak*

Pengenalan Suara Pembicara dengan Menggunakan Analisa MFCC Feature Extraction. Surabaya : Universitas Kristen Petra.

Muhammad Subekti : *Perbaikan Metode Backpropagation untuk Pelatihan Jaringan Syaraf Tiruan Multilayer,* Proceding Lokakarya Komputasi dan Sains Nuklir X, BATAN, 1999.

Nugroho Satriyo Anto, Witarto Budi Arief, Handoko Dwi.(2003). *Support Vector Machine Teori dan Aplikasinya dalam Bioinformatika*, pp 1-7.

Price, J. and Eydaghi, A. (2006). Design of Matlab-Based Automatic Speaker Recognition Systems. *9th International Conference on Engineering Education.* San Juan, PR.

Quatieri, Thomas F. (2001). *Discrete-Time Speech Signal Processing*. New Jersey: Prentice Hall.

Syah, D.P.A. (2009). *Sistem Biometriks Absensi Karyawan Dalam Menunjang Efektifitas Kinerja Perusahaan*. http://donupermana.wordpress.com/makalah/sistem-biometrik-absensi/. Akses tanggal : 23 Pebruari 2010.

Wangsa, G. and Gede, A.A. (2008). *Tugas Akhir: Sistem Identifikasi Telapak Tangan Dengan Menggunakan Metode Alihragam Fourier*. Bukit Jimbaran: Universitas Udayana.

Hartanto, B. 2008. *Memahami Visual C#.Net Secara Mudah*. Yogyakarta.

Yin Yin Aye (2009) Speech Recognition Using Zero-Crossing Features. *International Conference on Electronic Computer Technology*. Mandalay Technological University