REAL TIME SPEECH RECOGNITION SYSTEM

NOR RIFHAN BINTI ZAHARI

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

Faculty of Electronic and Computer Engineering
University Teknikal Malaysia Melaka

April 2009

© Universiti Teknikal Malaysia Melaka

UNIVERSTI TEKNIKAL MALAYSIA MELAKA

FAKULTI KEJURUTERAAN ELEKTRONIK DAN KEJURUTERAAN KOMPUTER

BORANG PENGESAHAN STATUS LAPORAN

PROJEK SARJANA MUDA II

Tajuk Projek : REAL TIME SPEECH RECOGNITION SYSTEM

Sesi Pengajian : 2008/2009

Saya NOR RIFHAN BINTI ZAHARI

(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)

Alamat Tetap: 1692 KAMPUNG KEPONG

20050 KUALA TERENGGANU

AMAT AMIR BIN BASARI
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: 29 APRIL 2009

Tarikh: 29 APRIL 2009

© Universiti Teknikal Malaysia Melaka

Special dedicated to my beloved parents, family, lectures, friends who had strongly encouraged
and supported me in my entire journey of learning

© Universiti Teknikal Malaysia Melaka

ACKNOWLEDGEMENT

I have completed my thesis which is partial fulfillment of requirements for the degree of Bachelor in Electronic Engineering (Computer Engineering).

On this opportunity, I would like to express my gratitude to the Faculty of Electronic Engineering, Universiti Teknikal Malaysia Melaka (UTeM) generally to the faculty and especially to my supervisor Mr. Amat Amir Basiri for help, advices and guidance throughout the process of searching, collecting information, analyzing and completing the report.

To my parents, I would like to express million of thanks to them for their support and love. Last but not least, I would like to thank my entire friend and also everyone who involve in this project either direct or indirectly.

© Universiti Teknikal Malaysia Melaka

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

Signature :...
Author : NOR RIFHAN BINTI ZAHARI
Date :.29 APRIL 2009...

© Universiti Teknikal Malaysia Melaka

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

Signature : Amat
Supervisor's Name : AMAT AMIR BIN BASARI
Date : 29 APRIL 2009

© Universiti Teknikal Malaysia Melaka

ABSTRACT

The purpose of this project is to develop a real time speech recognition system. The function of this speech recognition system is to the system unlock upon recognizing a voice password spoken by administrator or password holder. This project highlights the development of speech recognition using Matlab 7.0 to recognize the input speech. This system can be used as a security system such as Verification for Assessing Entry Application and Password Substitutions. The main objective of this project is to design the voice recognition system. Analysis of the voice is made using correlation and the regression method to compare the voiceprint of different words. These techniques provide stronger ability to recognize the same word. Voice training session is very important for the system to identify accurate frequency spectrum for one word. The training procedure is also used to reduce the random changes due to one word is spoken different times. The experimental results demonstrate high accuracy for this real time voice recognition system.

© Universiti Teknikal Malaysia Melaka

ABSTRAK

Projek ini bertujuan membina sebuah sistem pemgecaman suara secara terus (online).Sistem pengecaman suara ini dibina sebagai sistem keselamatan yang hanya berfungsi berdasarkan pengecaman pemilik suara sebagai kata laluan sesebuah sistem. Pembangunan sistem keselamatan ini dibina menggunakan perisian Matlab untuk proses pengecaman suara sebagai maklumat utama.Sistem ini boleh digunakan sebagai sistem keselamatan sebagai contoh Aplikasi pengecaman dan kemasukkan kata laluan. Isyarat sistem keselamatan akan berfungsi sepenuhnya apabila menerima ucapan atau pertuturan yang betul dan tepat. Objectif utama projek ini adalah untuk mereka bentuk sistem keselamatan menggunakan pengecaman suara. Pembina sistem ini haruslah mematuhi tata kerja sistem pengecaman suara untuk memastikan proses tersebut berjalan lancar. Kajian perlu dilakukan secara berkala untuk memastikan proses perbandingan suara dapat dilakukan dengan sempurna.Langkah kerja hendaklah sentiasa dilakukan bagi Mendapatkan frekuensi spektrum yang jitu. Selain itu, ralat paralaks dapat dielakkan. Dengan ini sistem pengecaman suara yang dilakukan untuk mendapatkan hasil yang sempurna adalah lebih tinggi.

© Universiti Teknikal Malaysia Melaka

© Universiti Teknikal Malaysia Melaka

© Universiti Teknikal Malaysia Melaka

© Universiti Teknikal Malaysia Melaka

LIST OF FIGURE

© Universiti Teknikal Malaysia Melaka

© Universiti Teknikal Malaysia Melaka

LIST OF TABLE

© Universiti Teknikal Malaysia Melaka

LIST OF ABBREVIATION

GUI	-	Graphical User Interface
CTT	-	The Centre for Speech Technology
ADC	-	Analog-to-Digital converter
MU	-	Motion Unit

© Universiti Teknikal Malaysia Melaka

CHAPTER 1

INTRODUCTION

1.1 Introduction

Biometric field research includes hand geometry, face prints, fingerprints, voiceprints, signatures, and non-retinal blood vessel analysis. Biometrics has been widely used in physical access control applications. Unlike personal identification number or pin, biometric features are something about the characteristic of a person. Biometric features are used to provide an enhanced level of security and identification. Pins and password may be forgotten and token-based identification method. Thus, biometric systems of identification are enjoying a new interest. Various types of biometric systems are being used for real-time identification. Speech recognition are the one of the most popular and reliable biometric features for verifying a person's identity. This function of speech recognition security system is to have a system that will only unlock upon recognizing of speech from password spoken by the administrator or password holder [2].

© Universiti Teknikal Malaysia Melaka

1.2 Objective

The objectives of this project are:

i) To by design the voice recognition system function using Matlab 7.0

ii) To recognize voice in real time function

1.3 Problem Statement

The speech recognition security system has been develop, comprised of speech recognition system that activated or unlock upon the security. The voice recognition system was capable of recognizing the password holder. Many of the application for identity authentication use a password or pin code. Other types of authentication such as signature, face and eye recognition are more complicated. The modern society has come to rely heavily on cards, passwords and pins when it comes to the safe guarding of resources and privacy, but as we all know these can sometimes be lost, stolen, cracked or simply forgotten, the very reasons why the world is moving towards the wide adoption of biometrics.

A small microphone was purchased and used to convert the human voice signal into a voltage signal. For information, speech technology offers some tangible advantages over alternative option if it to be successful in any given application.

1.4 Scope of Project

The scope of this project is mainly about the development of speech recognition for recognizing the voice of the same person by using the identification codes. The software is designed to detect the speech from the password holder and open when the password is correct. The software will be

© Universiti Teknikal Malaysia Melaka

controlled by Matlab. By using Matlab, it can be program by using MathWork tools.

1.5 Report Structure

This thesis consists of five chapters. Chapter I will describe about the brief overview and the definition about the project such as introduction, objectives, problem statement and scope of the project. This chapter there will be summary the project progress.

Chapter II will discuss about research and information which are related to this project. Every fact and information are gained from different references will be discussed so that the best technique and method can be implemented on this project. This will be based on the literature review and information about the project. Every facts and information which found through journals or other references will be compared and the better methods have been chosen for the project.

Chapter III will discuss t the project methodology used in this project such as data acquisition module, a pre-processing module, normalization and re-sampling module, a feature extraction module, a classifier module and a decision module. All these methodology should be followed for better performance.

Chapter IV will describe about the project finding such as progress result and analysis of the voice recognition. The result is presented by using tables, graph and figures.

The final chapter, Chapter V will explain about the conclusion of the whole project which includes project finding, achievement analysis and

© Universiti Teknikal Malaysia Melaka

conclusion about the research implementation which have been used. The project suggestion for enhancement also discussed.

© Universiti Teknikal Malaysia Melaka

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Basically this chapter will reveal the knowledge pertaining this field of project in which is gained through a lot of resources such as reference book, papers, journal, articles, conferences articles and documentations regarding applications and research work.

This shows how the theory and the concept have been implemented in order to solve project problem. The theory understanding is crucial as guidance to start any project. The result of the project cannot be assessed if it's not compared to the theory.

2.2 Background Of Study

Philippe Dreuw, David Rybach, Thomas Deselaers, Morteza Zahedi, and Hermann Ney ,produced a paper which tittle "Speech Recognition Techniques for a Sign Language Recognition System". The system developed is able to recognize sentences of continuous sign language independent of the speaker. The features used are obtained from standard video cameras without any special data acquisition devices. In particular, they

© Universiti Teknikal Malaysia Melaka

focus on feature and model combination techniques applied in ASR, and the usage of pronunciation and language models (LM) in sign language. These techniques can be used for all kind of sign language recognition systems and for many video analysis problems where the temporal context is important. For example; for action or gesture recognition. On a publicly available benchmark database consisting of 201 sentences and 3 signers, we can achieve a 17% WER.

They presented a vision-based approach to continuous automatic sign language recognition and have shown that appearance based features, which have been proven to be a powerful tool in many image recognition problems, are also well suited for the recognition of sign language. Furthermore, they have shown that many of the principles known from ASR, such as pronunciation and language modeling can directly be transferred to the new domain of vision based continuous ASLR. They presented very promising results on a publicly available benchmark database of several speakers which has been recorded without any special data acquisition tools. Combining different data sources, suitable language and pronunciation modeling, temporal contexts, and model combination, the 37% WER of our baseline system could be improved to 17.9% WER. The results suggest that for high dimensional data and the relatively low amount of available training data, PCA outperforms LDA for this task and that context information is as important as it is in ASR [12].

Shunji Mitsuyoshi, Fuji Ren, Yasuto Tanaka and Shingo Kuroiwa have written a paper with tittle of "Non-Verbal Voice Emotion Analysis System". A non-verbal voice analysis system that recognizes, separates and ranks concurrent emotions in real time has potential application in various fields, yet such a system that could delineate emotion based solely on the sound of a human voice has not been successfully demonstrated before. Here, they propose a system that recognizes human emotion by means of analyzing the fundamental frequency of a human voice taken from continuous natural speech. The system detects robust fundamental frequencies and intonations by parameterizing them into pitch, power, and deviation of power.

Based on these parameters, data was classified via decision-tree logic into the emotional elements of anger, joy, sorrow, and calmness. Degree of excitement was also

© Universiti Teknikal Malaysia Melaka

extracted. The system was evaluated by third parties by matching the system performance to human subjective classification for each element. Results indicate that overall matching rate was 70%, and the matching rate was 86% when compared to the subjects' assessment of their own voices. Their system performance exceeded the baseline with non-verbal information, which was equivalent to human subjective assessment [10].

Voice Recognition Security System was designed by Xiaowen Lu and Shihjia Lee in 2006. The function of this speech recognition security system is to have a system that will only unlock upon recognizing security system is to have a system that will only unlock upon recognizing a voice password spoken by the administrator or password holder. Matlab is used as platform to recognize the voice. Their design is based on the recorder program installed in Windows Xp and FFT function in Matlab. After we speak one word, a recorder program will stored the word in a.wav file. The microphone circuit goes to ADC of the MCU. The digitized sampling of the word is passed through the digital filters. The analysis is done on the MCU as well. Once that is done, the LCD which is connected to MCU displays if the word spoken matches the password or not [11].

Thai Automatic Speech Recognition was designed by Sinaporn Suebvisai1, Paisarn Charoenpornsawat, Alan Black, Monika Woszczyna, Tanja Schultz in 2005. They describe the development of a robust and flexible Thai Speech Recognizer as integrated into our English-Thai speech-to-speech translation system. They focus on the discussion of the rapid deployment of ASR for Thai under limited time and data resources, including rapid data collection issues, acoustic model bootstrap, and automatic generation of pronunciations. Issues relating to the translation and overall system will be reported elsewhere [13].

© Universiti Teknikal Malaysia Melaka

Real-Time Speech-Driven Face Animation With Expressions Using Neural Networks was designed by Pengyu Hong, Zhen Wen, and Thomas S. Huang in 2002. A real-time speech-driven synthetic talking face provides an effective multimodal communication interface in distributed collaboration environments. Nonverbal gestures such as facial expressions are important to human communication and should be considered by speech-driven face animation systems. In this paper, they present a framework that systematically addresses facial deformation modeling, automatic facial motion analysis, and real-time speech-driven face animation with expression using neural networks. Based on this framework, they learn a quantitative visual representation of the facial deformations, called the motion units (MUs). A facial deformation can be approximated by a linear combination of the MUs weighted by MU parameters (MUPs). They develop an MU-based facial motion tracking algorithm which is used to collect an audio–visual training database. Then, they construct a real-time audio-to-MUP mapping by training a set of neural networks using the collected audio–visual training database. The quantitative evaluation of the mapping shows the effectiveness of the proposed approach. Using the proposed method, they develop the functionality of real-time speech-driven face animation with expressions for the iFACE system. Experimental results show that the synthetic expressive talking face of the iFACE system is comparable with a real face in terms of the effectiveness of their influences on bimodal human emotion perception [15].

Speech Transcript Analysis for Automatic Search was designed by Anni R. Coden and Eric W. Brown in year 2001. They address the problem of finding collateral information pertinent to a live television broadcast in real time. The solution starts with a text transcript of the broadcast generated by an automatic speech recognition system. Speaker independent speech recognition technology, even when tailored for a broadcast scenario, generally produces transcripts with relatively low accuracy. Given this limitation, they have developed algorithms that can determine the essence of the broadcast from these transcripts. Specifically, we extract named entities, topics, and sentence types from the transcript and use them to automatically generate both structured

© Universiti Teknikal Malaysia Melaka

and unstructured search queries. A novel distance-ranking algorithm is used to select relevant information from the search results. The whole process is performed on-line and the query results (i.e., the collateral information) are added to the broadcast stream [16].

2.2 Overview On Speech Recognition

Today, a lot of researchers have carried out researchers on speech recognition which has become an active topic for biometrics fields. A wide range of methods has been tried out to make speech recognition research a success. Speech recognition is the process of taking the spoken word as an input to a computer program. This process is important to virtual reality because it provides a fairly natural and intuitive way of controlling the simulation while allowing the user's hands to remain free. This article will delve into the uses of voice recognition in the field of virtual reality, examine how speech recognition is accomplished, and list the academic disciplines that are central to the understanding and advancement of speech recognition technology [2].

Speech recognition comprised of two separate types of technologies. It is voice-scan and speech recognition. Voice-scan is used to authenticate a user based on user voice characteristics, while speech recognition is used for the technological comprehension of spoken words. Both play a role in voice recognition biometrics, and the science of virology is the underlying motivation [2].

© Universiti Teknikal Malaysia Melaka