SPEECH RECOGNITION USING RADIAL BASIS FUNCTION NEURAL NETWORK

NURUL SHAIRAH BT HASNA MERICAN

This report is submitted in partial fulfilment of requirement for the award of Bachelor Electronic Engineering (Industrial Electronic) with Honours.

Faculty of Electronic and Computer Engineering
Universiti Teknikal Malaysia Melaka

April 2010

UNIVERSTI TEKNIKAL MALAYSIA MELAKA
FAKULTI KEJURUTERAAN ELEKTRONIK DAN KEJURUTERAAN KOMPUTER

BORANG PENGESAHAN STATUS LAPORAN
PROJEK SARJANA MUDA II

Tajuk Projek	:	**Speech Recognition Using Radial Basis Function Neural Network (RBFNN)**
Sesi Pengajian	:	**0 9 / 1 0**

Saya NURUL SHAIRAH BT HASNA MERICAN
(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: 30/4/2010

Disahkan oleh:

(COP DAN TANDATANGAN PENYELIA)

KHAIRUL AZHA BIN A. AZIZ
Jurutera Pengajar
Fakulti Kejuruteraan Elektronik Dan Kejuruteraan Komputer
Universiti Teknikal Malaysia Melaka (UTeM)
Karung Berkunci No 1752
Pejabat Pos Durian Tunggal
76109 Durian Tunggal, Melaka

Tarikh: 30/4/2010

"I hereby declare that this thesis entitled, Speech Recognition Using Radial Basis Function Neural Network is the result of my own research idea concept for work that have been cited clearly in the references."

Signature : 

Name : NURUL SHAIRAH BT HASNA MERICAN

Date : 30/04/2010

"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 Electronic) with Honours"

Signature : .....

Supervisor's Name : EN. KHAIRUL AZHA B. ABD. AZIZ

Date : 30/04/2010

ACKNOWLEDGEMENTS

Alhamdulillah , I am most grateful to Almighty ALLAH s.w.t. for blessing me with good health and ideas in completing this research successfully.

Thousand of thanks to my supervisor, En. Khairul Azha b A. Aziz for giving me a chance to do my project under his supervises. I would like to show my highest gratitude for his invaluable support, patient, assistance and especially his encouragement to this project. I truly have learnt a lot and all this would not be without his guidance.

I also would to thanks to my family for their supporting especially my parents. Lots of love for them. To my siblings and family members who were always stand by my side to encourage, advice and support me during this entire project.

Last but not least, to all my fellow friends, thank you very much for your contribution in giving me a moral support throughout my project development period.

Lastly, I really appreciate to have this responsibility to finish this project. This task has taught a lot of lesson and knowledge which is much valuable for me in the future.

ABSTRACT

Speech recognition is the process of converting an acoustic signal, captured by a microphone or a telephone, to a set of words. The recognized words can be the final results, as for applications such as commands & control, data entry, and document preparation. They can also serve as the input to further linguistic processing in order to achieve speech understanding, a subject covered in section. In this paper, a speech recognition system using neural network (NN) with Radial Basis Function Neural Network (RBFNN) method is proposed. The training speed of RBFNN can be orders of magnitude faster [1] than the well known back propagation paradigm, and yet the ability of the network to generalize to detect the voice is approximately the same [2]. From this project, RBFNN method is to test the main and other voices. Then, show the output that and tell that the system can detect.

ABSTRAK

Pengecaman pertuturan ialah proses menukar satu isyarat akustik, ditangkap oleh sebuah mikrofon atau sebuah telefon, untuk satu set kata-kata. Kata-kata sedar boleh jadi keputusan muktamad, bagi permohonan-permohonan seperti menguasai & kawalan, kemasukan data, dan penyediaan dokumen. Mereka boleh juga digunakan sebagai input untuk pemprosesan yang seterusnya linguistik supaya mencapai pemahaman pertuturan, satu subjek dilindungi dalam seksyen. Dalam kertas ini, sistem pengecaman pertuturan menggunakan jaringan saraf (NN) dengan jaringan saraf fungsi asas jejari (RBFNN) kaedah adalah dicadangkan. Kelajuan latihan RBFNN boleh jadi urutan magnitud lebih cepat [1] daripada paradigma penyebaran belakang terkenal, tetapi keupayaan rangkaian untuk menggeneralisasi bagi mengesan suara ialah kurang sama banyak [2]. Daripada projek ini, kaedah RBFNN ialah untuk menguji utama dan suara-suara lain. Kemudian, menunjukkan keluaran yang dan memberitahu yang sistem itu dapat mengesan.

CONTENT

LIST OF FIGURE

LIST OF ABBREVIATION

NN	-	Neural Network
ASR	-	Automatic Speech RecognitionProbabilistic Neural Network
RBFNN	-	Radial Basis Function Neural Network

CHAPTER 1

INTRODUCTION

This chapter will briefly discuss on the project overview. The objective, scope, problem statement, summary of methodology and thesis outline will be presented in this chapter.

1.1 Project Background

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 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[12].

1.2 Objective

The objective of this project are :

- To design the voice recognition system by using Matlab's software.
- Developed by using MATLAB's software.

1.3 Problem Statement

The speech recognition 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 identify 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 vary reason 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. Some problems can be solve by using this system are:

- How to make it easier for people to access function commands, such as opening files and accessing menus with voice instruction?
- How can this system will help people with disabilities that prevent them from typing have also adopted speech recognition system?

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 speech. One voice is selected as the reference or the main speech for the recognition. The software is designed to detect the speech. Limit to 20 inputs to train the data and by using the figure of the voice recorded to process the system. The software will be controlled by Matlab. By using Matlab, it can be program by using the coding referring to the MathWork tools and other references book about Matlab.

1.5 Report Structure

This thesis consists of five chapters as well. For Chapter 1 will describe about the brief overview and the definition about the project such as introduction, objectives, problem statement and scope of the project.

Meanwhile chapter two discuss about the background of study related to speech recognition using RBFNN method. Literature review will produce overall structure of the face detection which shows the relationship between project research and theoretical concept.

Chapter three will explain about the project methodology. Project methodology give details about the method used to solve the problem to complete the project. The method used such as collecting data method, process and analysis data method, modelling and etc.

Chapter four consists of result and discussion of the project, finding and analysis throughout the research and project development.

Lastly, chapter five is the project conclusion. This chapter rounds up the final achievement of the whole project and reserves suggestions for possible future researches.

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, journals, articles, conferences articles and documentations regarding applications and research work.

This shows how the theory and 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 is not compared to the theory.

2.2 Background of Study

What is a speech recognition. The term "voice recognition" is sometimes used to refer to speech recognition where the recognition system is trained to a particular speaker. As is the case for most desktop recognition software, hence there is an aspect of speaker recognition, which attempts to identify the person speaking, to better

recognize what is being said. Speech recognition is a broad term which means it can recognize almost anybody speech, such as a call centre system designed to recognize many voices. Voice recognition is a system trained to a particular user, where it recognize their speech based on their unique vocal sound.

Speech recognition is one of the desired assistive technology systems. People believe speech recognition is a natural and easy method of accessing the computer but there are some difficulties to generate it according to those problems below :

- training,
- microphone selection,
- cognitive abilities,
- environmental area,
- consistent speech.

Also, speech recognition software is used in homes and businesses. A range of software products allows users to dictate to their computer and have their words converted to text in a word processing or e-mail document. You can access function commands, such as opening files and accessing menus, with voice instructions. Some programs are for specific business settings, such as medical or legal transcription. People with disabilities that prevent them from typing have also adopted speech-recognition systems. If a user has lost the use of his hands, or for visually impaired users when it is not possible or convenient to use a Braille keyboard, the systems allow personal expression through dictation as well as control of many computer tasks. Some programs save users' speech data after every session, allowing people with progressive speech deteriorations to continue to dictate to their computers.

Philippe Dreuw, David Rybach, Thomas Deselaers, Morteza Zahedi, and Herman Ney, produced a paper which is title by "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 focused on feature and model combination techniques applied in Automatic Speech Recognition, and the usage of pronunciation and Language Models 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.

They presented a speech recognition 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. Further more, they have shown that many of the principles known from Automatic Speech Recognition, such as pronunciation and language modeling can directly be transferred to the new domain of vision based continuous Automatic Speech Language Recognition. 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 and that context information is as important as it is in Automatic Speech Recognition [12].

Thai Automatic Speech Recognition was designed by Sinaporn Suebvisail, Paisarn Charoenpornsawat, Alan Black, Monika Woszcyna, 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].

2.3 Overview On Speech Recognition

Today, a lot of researchers have carried out on speech recognition which has become an active topic for biometrics fields. A wide range of methods have 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 simulation while allowing the user's hand 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 for the technological comprehension of spoken words. Both play a role in voice recognition biometrics, and the science of virology is the underlying motivation.

2.4 The Speaking Voice and Factors to Consider

Analyzing sound waves can produce a series of voice patterns that are based on frequency, intensity and time, among other factors. The voice involves the anatomy of larynx and the elements of their voices are affected by its components, in addition to any potential disorders or larynx is positioned in the anterior neck, slightly below the point where the pharynx divides gives rise to the separate respiratory and digestive tracts[2].

2.5 Voice analysis

2.5.1 Wideband Spectograms

Wideband spectrograms are marked by the relatively broad bands of energy that depict the formants. The centre of each band of energy is taken to be the frequency of the formant, and the range of frequencies occupied by the band is taken to be the bandwidth of the formant. The relative degree of darkness of a band of energy can be used as a rough estimate of the intensity of the signal, and the relatively large horizontal blank spaces between the formants represent troughs (antiformants) in the resonance

curve of the vocal tract. Information about the timing of changes in vocal tract resonance is more reliably obtained from wideband spectrograms. Unlike a narrowband spectrogram, a wideband display will effectively represent an aperiodic source that is being resonant in the vocal tract. The bandwidth of a filter used to generate a wideband spectrogram is generally between 300 and 500 Hz. A filter with such a relatively wide bandwidth will respond in the same way to one, two, three or even more harmonics that fall within its range: the filter will not resolve the energy within its bandwidth into individual harmonics[5].

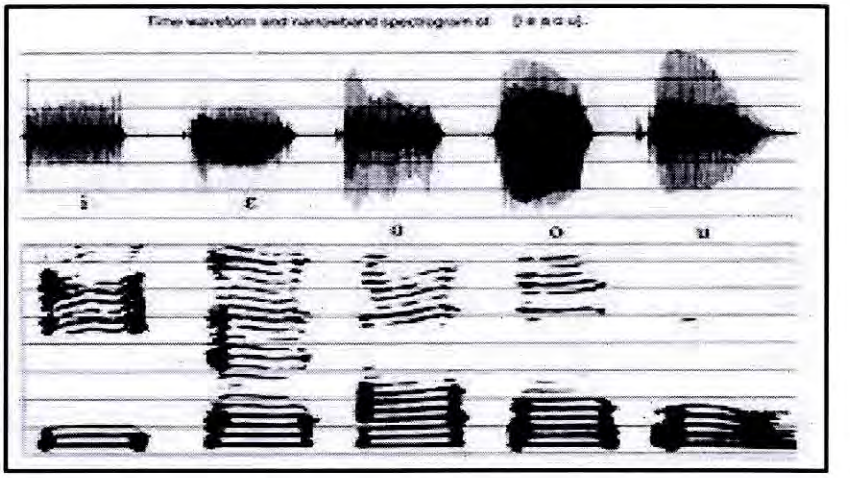

Figure 2.1 : Wideband Spectograms

2.5.2 Narrowband Spectograms

Narrowband spectrograms are marked by the more or less narrow horizontal bands which represent the harmonics of the glottal source. The darker bands represent the harmonics that are closest to peaks of resonance in the vocal tract. The lighter bands represent harmonics whose frequencies are further away from the resonance peaks. The bandwidth of the filter used to generate narrowband spectrograms is usually between 30 and 50 Hz. As the fundamental frequency is unlikely to be lower than 50 Hz, a filter with that bandwidth will respond to and capture each harmonic separately as it scans

through the frequencies in the speech signal. Narrowband spectrograms have traditionally been used for making measurements of fundamental frequency and intonation[5].

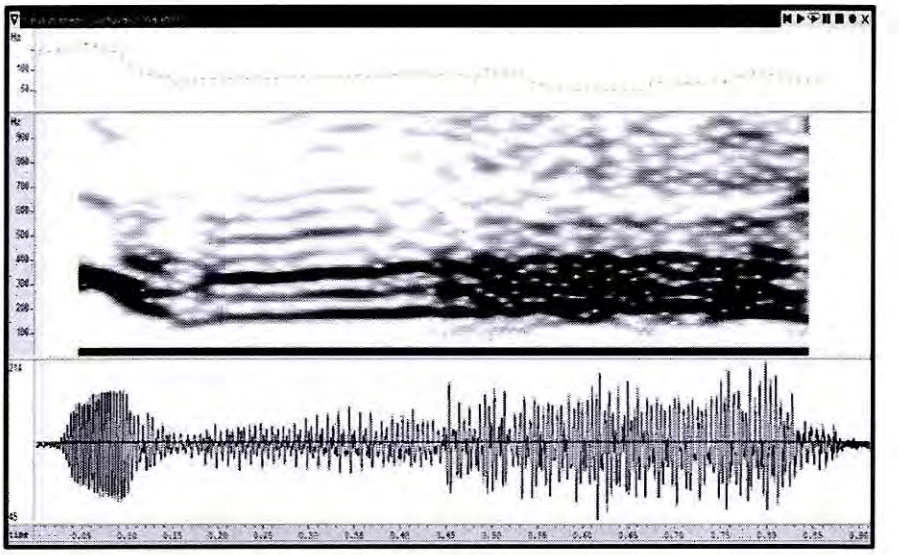

Figure 2.2 : Narrowband Spectograms

2.6 Radial Basis Function Neural Network

A Radial Basis Function Neural Network has an input layer, a hidden layer and an output layer. The neurons in the hidden layer contain Gaussian transfer functions whose outputs are inversely proportional to the distance from the center of the neuron. In order to use radial basis function, we need to specify the hidden unit activation function, the number of processing units, a criterion for modeling a given task and a training algorithm for finding the parameters of the network. Finding the RBF weight is called network training. If we have at a set of input-output pairs, called training set, we optimize the network parameters in order to fit the network output to the given inputs. The fit is evaluated by means of a cost function, usually assumed to be the mean square error. After training, the RBF network can be used with data whose underlying statistics is similar to that of the training set.

2.6.1 Radial Basis Function

Different types of radial basis functions could be used, but the most common is the Gaussian function:

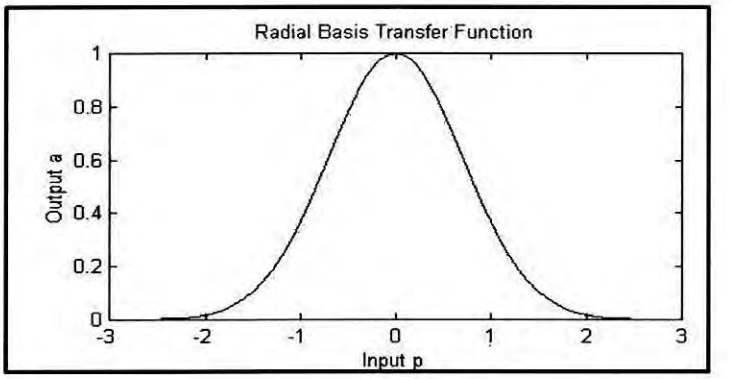

Figure 2.3 : Radial Basis Transfer Function

If there is more than one predictor variable, then the RBF function has as many dimensions as there are variables. The following picture illustrates three neurons in a space with two predictor variables, X and Y. Z is the value coming out of the RBF functions:

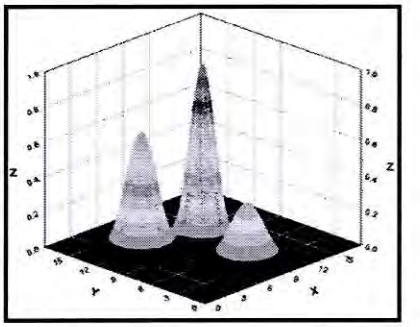

Figure 2.4 : Three neurons in a space X,Y and Z.