SPEAKER IDENTIFICATION USING WAVELET PACKET TRANSFORM AND FEED FORWARD NEURAL NETWORK

By

MOHAMED ALI ALMASHRGY

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Master Science

March 2005

DEDICATION

TO

MY BELOVED FAMILY

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfillment of the requirement for the degree of Master of Science

SPEAKER IDENTIFICATION USING WAVELET PACKET TRANSFORM AND FEED FORWARD NEURAL NETWORK

By

MOHAMED ALI ALMASHRGY

March 2004

Chairman: Associate Professor Adznan Bin Jantan, Ph.D.

Faculty: Engineering

It has been known for a long time that speakers can be identified from their voices. In this work we introduce a speaker identification system using wavelet packet transform. This is one of a wavelet transform analysis for feature extraction and a neural network for classification. This system is applied on ten speakers

Instead of applying framing on the signal, the wavelet packet transform is applied on the whole range of the signal. This reduces the calculation time. The speech signal is decomposed into 24 sub bands, according to Mel-scale frequency. Then, for each of these bands, the log energy is taken. Finally, the discrete cosine transform is applied on these bands. These are taken as features for identifying the speaker among many speakers.

For the classification task, Feed Forward multi layer perceptron, trained by backpropagation, is proposed for use as training and classification feature vectors of the speaker. We propose to construct a single neural network for each speaker of interest.

Training and testing of isolated words in three cases, Vis one-, two-, and three-syllable words, were obtained by recording these words from the LAB colleagues using a low-cost microphone.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhli keperluan untuk ijazah Master Sains

PENGENALPASTIAN PERTUTURAN MENGGANKAN KAEDAH MENGUBAH PAKET DAN SUAPAN HADAPAN RANGKAIAN NEURAL

Oleh

MOHAMED ALI ALMASHRGY

March 2004

Pengerusi: **Profesor Madya Adznan Bin Jantan, PhD**

Fakulti: **Kujuruteraan**

Sebagaimana telah diketahui sejak dari dulu lagi, seseorang boleh dikenalpasti berdasarkan suaranya. Tesis ini, memperkenalkan sistem pengecaman seseorang yang bercakap dengan menggunakan transformasi paket gelombang, dengang menggunakan analisis transformasi gelombang untuk penguraian sifat gelombang dan rangkaian neural untuk klasifikasi. Sistem ini telah diimplementasikan pada sepuluh orang.

Sistem ini menggunakan transformasi paket gelombang pada keseluruhan had gelombang, bagi menggantikan penggunaan merangka gelombang. Ini dapat mengurangkan masa pengiraan. Di dalam sistem ini, gelombang ucapan telah diurai menjadi 24 sub-lingkaran yang berdasarkan skala frekuensi Mel. Selepas itu, tenaga log bagi setiap lingkaran diambilkira. Akhir sekali, transformasi diskrit kosinediimplementasikan pada lingkaran-lingkaran tersebut. Ianya diambilkira sebagai sifat-sifat untuk mengenalpasti orang yang bercakap di antara orang-orang lain.

Bagi tugas mengklasifikasi, Feed Forward Multi Layer Perceptron, yang telah dilatih oleh backpropagation, digunakan sebagai kaedah latihan dan klasifikasi sifat-sifat vektor orang yang bercakap. Tesis ini, mencadangkan penggunaan satu rangkaian neural bagi setiap orang yang bercakap.

Perkataan-perkataan yang digunakan untuk latihan dan ujian rangkaian neural telah dibahagikan kepada tiga kes; satu-, dua-, and tiga- patah perkataan. Ianya telah diperolehi dengan merakamkan perkataan yang disebut oleh rakan-rakan di Makmal Multimedia dan Pemprosesan Imej dengan menggunakan mikrofon yang murah.

ACKNOWLEDGEMENTS

The first person I would like to thank is my supervisor Assoc. Prof. Dr. Adzna Bin Jantan who kept an eye on the progress of my work and always was available when I needed his advises. I owe him lots of gratitude for having me shown this way of research. Besides of being an excellent supervisor, Dr. Adznan was as close as a relative and a good friend to me.

I would also like to thank the other members of my M. Sc. committee who monitored my work and took effort in reading and providing me with valuable comments on earlier versions of this thesis: Assoc. Prof. Dr. Abd Rahman Ramli, and Dr. Elsadig Ahmed Mohamed Babiker.

I feel a deep sense of gratitude for my late father and mother who formed part of my vision and taught me the good things that really matter in life. The happy memory of my father still provides a persistent inspiration for my journey in this life. I am grateful for my brother Abdul Gadeer, and my sisters, for rendering me the sense and the value of brotherhood. I am glad to be one of them. The chain of my gratitude would be definitely incomplete if I would forget to thank my uncle Mohamed. Also, I am very grateful for my fiancée, for her love and patience during the M. Sc period.

Also, lots of thanks go to all of my friends in Malaysia, and Libya for their engorgements during my study.

I certify that an Examination Committee met on 17/3/2004 to conduct the final examination of Mohamed Ali Almashrgy on his Master of Science thesis entitled "Speaker Identification System Using Wavelet Packet Transform and Feed Forward Neural Network" in accordance with Universiti Pertanian Malaysia (Higher Degree) Act 1980 and Universiti Pertanian Malaysia (Higher Degree) Regulations 1981. the committee recommended that the candidate be awarded the relevant degree. Member of the Examination Committee are as follow:

Chairman, PhD
Professor
Name of faculty/institute
Universiti Putra Malaysia
(Chairman)

Examiner 1, PhD
Professor
Name of faculty/institute
Universiti Putra Malaysia
(Member)

Examiner 2, Ph.D.
Professor
Name of faculty/institute
Universiti Putra Malaysia
(Member)

Independent Examiner, Ph.D.
Professor
Name of faculty/institute
Universiti Putra Malaysia
(Independent Examiner)

_____
GULAM RUSUL RAHMAT ALI, Ph.D.
Professor/Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date :

This thesis submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfillment of the requirements for the degree of Master of Science. The members of the Supervisory Committee are as follows:

Adznan Bin Jantan, PhD
Associate Professor
Faculty of Engineering
University Putra Malaysia
(Chairman)

Abd Rahman Ramli, PhD
Associate Professor
Faculty of Engineering
University Putra Malaysia
(Member)

Elsadig Ahmed Mohamed Babiker, PhD
Lecture
Faculty of Engineering
University Putra Malaysia
(Member)

_____

AINI IDERIS, PhD
Professor/Dean
School of Graduate Studies
Universiti Putra Malaysia

Date :

DECLARATION

I hereby declare that the thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at UPM or other institutions.

_____

MOHAMED ALI ALMASHRGY

Date:

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABREVIATION

WPT	Wavelet Packet Transform
DWT	Discrete Wavelet Transform
NN	Neural Network
ZCR	Zero Crossing Rate
LPC	Linear Prediction Coding
FFT	Fast Fourier transform
DTW	Dynamic time warping
VQ	Vector Quantization
HMM	Hidden Markov Model
GMM	Gaussian Mixture Model
STFT	Short Term Fourier Transform
WFT	Windowed Fourier Transforms
MLP	Multi-Layer-Perceptron
LMS	Linear recursive least-mean-square
ART	Adaptive Resonance Theory
FFE	Further Feature Extract
LDA	Linear Discriminate Analysis
LPCC	Linear Predictive Cepstral Coefficients
FFNN	Feed Forward Neural Network
DCT	Discrete Cosine Transform
ITU	Upper energy threshold
ITL	Lower energy threshold

IZCT	Zero crossings rate threshold
EPD	EndPoint Detection
MLFF	Multi Layer Neural Network

CHAPTER 1

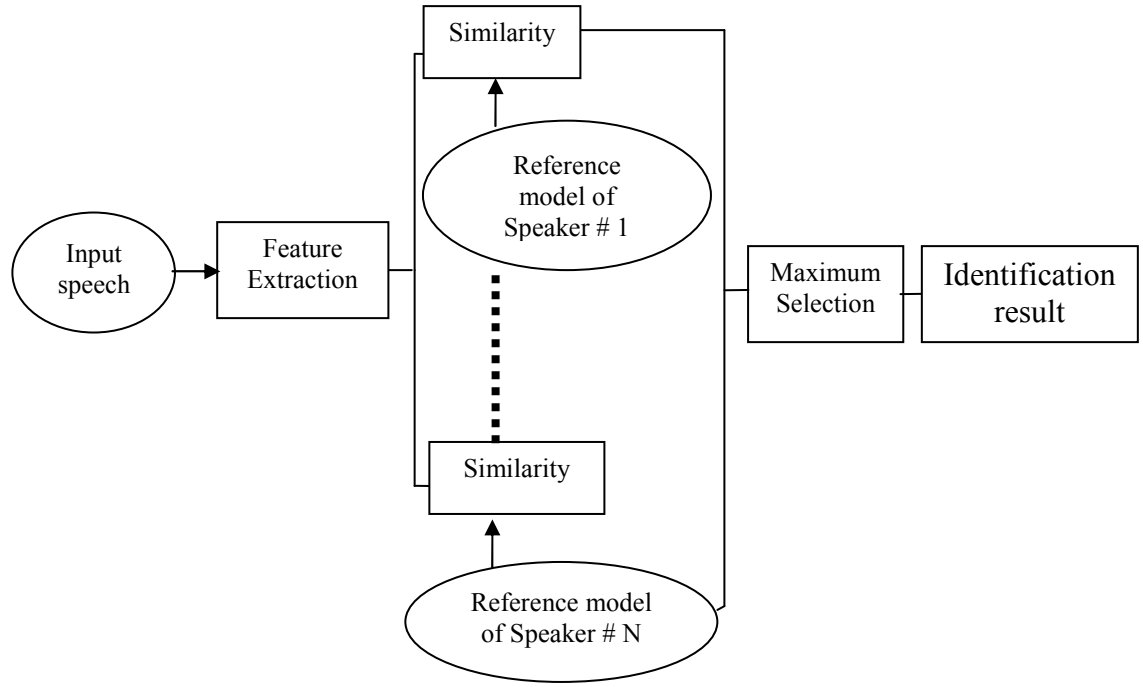
INTRODUCTION

1.1 Background

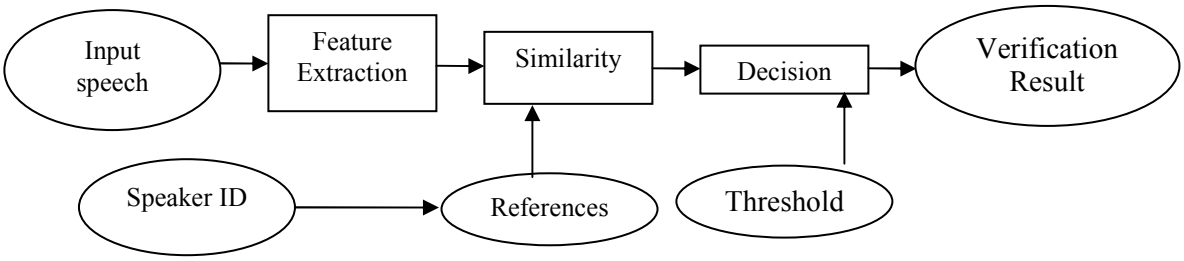
The speech signal conveys many level of information to the listener. At the primary level, speech conveys a message via words. But other levels speech conveys information about the language being spoken and the emotion, gender and, generally, the identity of the speaker. While speech recognition aims at recognition system is to extract, characterize and recognize the information in the speech signal conveying speaker identity (D.A. Reynolds, 2002).

Figure 1.1 shows a simple diagram of speaker identification and verification system.

Speaker recognition is a generic term for the classification of a speaker's identity from an acoustic signal. In the case of speaker identification, the speaker is classified as being one of the finite sets of speakers. As in the case of speech recognition, this will require the comparison of a speech utterance with a set of references for each potential speaker. For the case of speaker verification, the speaker is classified as having the claimed identity or not. That is, the goal is to automatically accept or reject an identity that is claimed by the speaker. In this case, the user will first identify herself or himself, and the



(a) Speaker Identification



(b) Speaker Verification

Figure 1.1 The speaker identification and verification system.

distance between the associated reference and the pronounced utterance will be compared to threshold that is determined during training. Speaker recognition based on text-dependent and text-independent utterance. The former require the speaker to say key words or sentences having the same text for both training and recognition trials, whereas the latter do not rely on a specific text being spoken (D.A. Reynolds, 2002).

Both text-dependent and independent methods share a problem however. These systems can be easily deceived because someone who plays back the recorded voice of a registered speaker saying the key words or sentences can be accepted as the registered speaker (Cole, R.A et al, 1996).

1.2 Speaker recognition application

The speaker recognition has many potential applications. The applications of speaker recognition technology are widely used, continually growing. The applications of speaker recognition are beyond of scope our research. Therefore, the following section gives a short discussion for these applications:

1.2.1 Access control: Used for controlling access to computer networks or web sites. Also used for automated password (D.A. Reynolds, 2002).

1.2.2 Transaction authentication: For telephone banking, in addition to account access control, high levels of verification can be used for more sensitive transactions. More recent applications are in user verification for remote electronic and mobile purchase (e- and commerce) (D.A. Reynolds, 2002).

1.2.3Speech data management: In voice mail browsing or intelligent answering machine, use speaker recognition to label incoming voice mail with speaker name for browsing and/or for action (D.A. Reynolds, 2002).

1.3 Scope of work

In this thesis the wavelet packet transforms proposed to use as speaker recognition due to their benefits compared with STFT. The most important thing in wavelet packet transform is the localization in frequency. The MATLAB 6.5 platform has been used for our text dependant speaker recognition system. A number of speakers used for testing the system. The features of some of them are extracted and stored which used for classification. And the other considered as an imposter to calculate the accuracy of the system. Each speaker record his speech signal of English words at one, two, and three syllable words for training and testing set. The test data set used for testing the system are different from the one used for the training. The system will identify whether the speaker who has spoken to the system is from the allowed speaker to enter or not. We use database for training and testing. The sampling frequency for the database is 16000 HZ.

1.4 Problem statements

The main problem in speaker recognition, as with another complex task that require some form of intelligence, is the amount of information that must be examined before making a classification (M. Mills, 1996). The main idea behind the speaker recognition system is to extract some features from the speech signal of the speaker, which will be used as a reference to replace the signal itself. This extraction called Front-end processing.

One of the new ideas that have for the past decade been extensively studied for different applications is that of wavelet transform. This transform has gained a great admires in many fields including signal processing. Signal processing researchers have adopted various types of wavelet analysis methods which reports improvement over traditional techniques. Speech signal processing area is one of those areas in which gain great improvements.

There were some of speech signal processing researches such as speech recognition, speaker recognition and speech compression where the wavelet transforms is applied.

The Discrete Wavelet Transform (DWT) is one of the wavelet transforms was used by (Siew et al, 2001) in the area of speaker recognition. The results indicated that DWT could be a potential features extraction tool for speaker recognition.

In this thesis, wavelet packet transform (one of the wavelet transforms) is applied for extracting features in the area of speaker identification task instead of using DWT.

In this thesis, the wavelet packet transform is proposed to use for extracting a features of speakers from their speech signal. This is due to the advantages of using WPT that it can segment the frequency axis and has uniform translation in time.

On the other hand, Feed Forward multi layer perceptron trained by back propagation is used to train and classify the feature vectors of the speaker. Instead of using one NN for all the speakers, each speaker has it is own NN.

1.5 The Objectives

1. Study different methods for analysis speech signal and further discuss the wavelet packet transform.
2. Study different methods for classification and further discuss on neural network which is used in this thesis.
3. Design a speaker identification system using the proposed system.
4. Implement this system on text dependent speaker recognition.
5. Evaluation of the system's performance.

1.6 Thesis organization

- Chapter 1

This chapter explains a brief background to the area of speaker recognition, and the application of speaker recognition. After that, explain the scope of work, the problem statement and finally the thesis organization.

- Chapter 2

Chapter 2 provides an overview of the basic techniques for speech recognition of speaker recognition, proposed in the literature over the past

few years. The two main issues relevant to the design speaker recognition systems are: First, the front end processing (feature extraction) that captures speech feature and, second, the classification of features. Both are challenging problems, and significant research effort has been directed towards finding appropriate solutions.

- Chapter 3

This chapter covers the methods, which used for this system. Due to the concentration in this thesis for extracting feature from speaker so much work has been done for explaining the wavelet transform.

- Chapter 4

Chapter 4 presents the results which are obtained using our proposed method for speaker identification system.

- Chapter 5:

This chapter contains the conclusion and suggestion for future work.