



Physics Department
Faculty of Science
Ain Shams University

Design and Implementation of a Stand-alone Voice Recognition System

A Thesis

**“Submitted for Partial Fulfillment of the Requirements for
M.Sc. Degree in Physics”**

By

Mohammed Ramadan Saady Abd El Tawab

Demonstrator at Basic Science Department,
Faculty of Computers and Information,
Fayoum University,
Cairo, Egypt.

Supervised by

1- Prof. Dr. Ashraf shamseldin Yahia

Professor of Electronics and Electromagnetics,
Physics Department, Faculty of Science,
Ain Shams University.

2- Emeritus Prof. Dr. Hatem M. El Borai

Professor of Electronics and Circuits,
Physics Department, Faculty of Science,
Ain Shams University.

3- Dr. El-Sayed A. El-Dahshan

Assistant Professor of Computational physics,
Physics Department, Faculty of Science,
Ain Shams University.

2016



Physics Department
Faculty of Science
Ain Shams University

Approval sheet

Name: Mohammed Ramadan Saady Abd El Tawab

Title: Design and Implementation of a Stand-alone
Voice Recognition System

Supervisors

Signature

Prof. Dr. Ashraf shamseldin Yahia

.....

Professor of Electronics and Electromagnetics,
Physics Department, Faculty of Science,
Ain Shams University.

Emeritus Prof. Dr. Hatem M. El Borai

.....

Professor of Electronics and Circuits,
Physics Department, Faculty of Science,
Ain Shams University.

Dr. El-Sayed A. El-Dahshan

.....

Assistant Professor of Computational physics,
Physics Department, Faculty of Science,
Ain Shams University.

Acknowledgements

First and foremost praising to my god, Allah who led me through this work sustains me with the power and will to do this work.

My very deep gratitude goes to Prof. Dr. **Salah Yaseen El-Bakry**, chairman of the physics department, for his interest, fruitful comments, support and encouragement.

I would like to express my deep gratitude to Prof. Dr. **Ashraf Shamseldin Yahia** for his supervision, guidance and continuous support to me, during the study period and his handling upscale and humble, which had a major impact in the completion of this work.

I am profoundly grateful to Assoc. Prof. Dr. **El-Sayed A. El-Dahshan** for suggesting this work, and for his help to overcome all the difficulties and obstacles that I faced during the work.

I would like to express my deep gratitude to Emeritus Prof. Dr. **Hatem El-Borey** for his supervision, great efforts, fruitful discussions, and continuous help during the preparation of this thesis.

I want to express a lot of thanks to my Prof. Dr. **Nabila Mohammed Hassan** for her encouragement and support.

On a personal basis, my heartfelt thanks to my mother and my father who are always doing supplication and prayers for me. Also I would like to extend my thanks to my wife for her support and encouragement. My thanks go to my brothers and my sister for their encouragement.

Abstract

For the past several decades, the biometric security systems varied between those based on behavioral and others based on physiological features of people. One of the behavioral biometric security systems is that based on voice features of people and is called speaker recognition system. As a result of the advance of machine learning and computer technology, speaker recognition has rapidly evolved and has become very popular in the recent years. It is being intensively researched and found many applications where it saves lot of troubles which appear as a result of using other biometric security systems, so it is considered a high security system.

Many methods were presented for the aim of designing an automatic speaker recognition system and other methods were presented for the aim of developing reliable speaker recognition systems. In general, any speaker recognition system involves four basic steps: a) data base formation, b) pre-processing, c) features extraction, and d) classification or matching of extracted features.

In this work we have developed a method for speaker recognition using the English Language Speech Database for Speaker Recognition (ELSDSR) database which is compose of audio files for training and others for testing. The developed

method starts by pre-processing the training audio files of the database. Then the Wavelet Packet Transform (WPT) is employed on the pre-processed files for feature extraction purposes. For the excessive number of features provided with the WPT, the energy corresponding to each WPT node is calculated to reduce the dimensionality of the wavelet coefficients by removing redundant features and to form features vectors. The features vectors are sent to the Feed Forward Back-propagation Neural Network (FFBPNN) system.

Experimental results showed the effectiveness of the developed method by using the test audio files. Our results have also showed that the rate of correct recognition of the developed method is about 100% when using the training files and 95.7% when using a one testing file for each speaker from the ELSDSR database. The proposed method showed efficiency better than the well-known Mel Frequency Cepstral Coefficient (MFCC) and the Zak transform.

Table of Contents

Acknowledgement	I
Abstract	III
Table of Contents	V
List of Figures	X
List of Tables	XII
List of Abbreviations	XIII
Chapter 1 Introduction	1
1.1 Literature Review.....	3
1.2 Motivation.....	6
1.3 Problem Statement.....	9
1.4 Thesis Contribution.....	10
1.5 Thesis Outline.....	12
Chapter 2 Speaker Recognition	14
2.1 Overview of Speaker Recognition.....	14
2.2 Speaker Verification System.....	16
2.3 Speaker Identification.....	18
2.4 Voice Production.....	19
2.4.1 Voice Production Unit.....	19
2.4.1.1 Excitation Production.....	21
2.4.1.2 Vocal Tract Articulation.....	23
2.5 The Human Recognition System.....	23
2.5.1 The Natural Auditory System.....	24

2.5.1.1 The Ear	24
2.5.2 The brain.....	25
2.6 Generic Speaker Recognition System	26
2.6.1 Pre-processing	27
2.6.2 Feature Extraction	27
2.6.3 Modeling.....	27
2.6.4 Matching.....	28
2.7 Measuring of efficiency of SRS	28
2.8 Speaker Recognition Applications	29
2.9 Current used Application Example	30
2.10 Most Common Techniques for Speaker Recognition .	31
2.10.1 Feature Extraction Techniques	31
2.10.2 Feature Matching Techniques.....	34
Chapter 3 Signal Pre-processing.....	35
3.1 Introduction	35
3.2 Signal Pre-processing	35
3.2.1 Digitalization.....	36
3.2.2 Truncation	37
3.2.3 Windowing	38
3.2.3.1 Convolution.....	38
3.2.3.2 Correlation	39
3.2.4 Energy Normalization	41
Chapter 4 Wavelet Transform.....	43
4.1 Introduction	43

4.2 Historical Development	44
4.3 Fourier Analysis	46
4.3.1 Fourier Transform.....	48
4.3.2 Fourier Series.....	49
4.3.3 Discrete Fourier Transform	49
4.3.4 Short Time Fourier Transform	51
4.4 What are Wavelets?	52
4.5 Wavelet Transform	55
4.5.1 The Continuous Wavelet Transform	56
4.5.2 The Discrete Wavelet Transform.....	58
4.5.2.1 Multi-Resolution Analysis	60
4.5.3 Wavelet Packet Transform	61
4.6 Similarities to Fourier Transform	63
4.7 Some of Most Popular Wavelet Types	63
4.7.1 Daubechies Wavelet.....	64
4.7.2 Mexican Hat Wavelet	64
4.7.3 Morlet Wavelet	65
4.8 Examples of Wavelet Filters	65
4.9 Choosing the Best Wavelet	67
4.10 Wavelet Applications	68
Chapter 5 Artificial Neural Network.....	70
5.1 Introduction	70
5.2 What are Artificial Neural Networks (ANNs)	70
5.3 Characteristics of Artificial Neural Networks	72

5.4 History and Development of Artificial Neural Networks	73
5.5 Biological Neuron	75
5.6 The Artificial Neuron.....	77
5.6.1 The Neuron Model	77
5.6.2 Some Activation Functions	79
5.7 Artificial Neural Networks Architecture	82
5.7.1 Feed-forward Networks	82
5.7.1.1 <i>Single Layer Feed-forward Network</i>	82
5.7.1.2 <i>Multi-layer Network</i>	83
5.7.2 Recurrent Networks	84
5.8 Learning Process	85
5.8.1 Learning Paradigms (models)	86
5.8.1.1 <i>Supervised</i>	86
5.8.1.2 <i>Non supervised</i>	87
5.8.2 Some Learning Algorithms	87
5.8.2.1 <i>Back-propagation Algorithm</i>	87
5.8.3 Learning Rules	89
5.8.3.1 <i>Error-Correction Rule</i>	89
5.8.3.2 <i>Hebbian Rule</i>	90
5.8.3.3 <i>Competitive Rule</i>	92
5.9 Artificial Neural Networks Applications	93
Chapter 6 Developed System and Results.....	95
6.1 Introduction	95
6.2 The Proposed Speaker Recognition System	95

6.2.1 The Database Used in This Study	96
6.2.2 Pre-processing of the Speech Signals	97
6.2.3 Feature Extraction Based on Wavelet Packet Transform.....	100
6.2.4 Classification by Artificial Neural Networks	102
Chapter 7 Hardware implementation	110
7.1 Introduction	110
7.2 Introducing the Speech Recognition Module	111
7.3 EasyVR2.0 Specifications.....	112
7.4 System Setup	113
7.4.1 Arduino Board Specifications	114
7.5 System Operation	115
7.6 Results	116
Chapter 8 Conclusion and Future Work.....	117
References	119

List of Figures

Figure No.	Title	Page No.
Figure 2-1	Speech processing and its applications	14
Figure 2-2	Speaker Verification system	17
Figure 2-3	Speaker Identification system	18
Figure 2-4	Voice Production Unit	21
Figure 2-5	Anatomy of the human ear	25
Figure 2-6	Generic Speaker Recognition System	26
Figure 3-1	The convolution strategy	39
Figure 3-2	The correlation strategy	41
Figure 4-1	Decomposition of 16 point time domain signal by FFT through four steps	50
Figure 4-2	STFT of the signal and the time-frequency representation.	52
Figure 4-3	Decomposition a signal by using wavelet transform	53
Figure 4-4	(a) a continuous cosine wave, (b) a continuous wavelet	54
Figure 4-5	Examples of various wavelet types	55
Figure 4-6	Steps of correlating a CWT to a signal	58
Figure 4-7	Single level DWT	60
Figure 4-8	The 2-level WPT	62
Figure 4-9	Mexican Hat wavelet	64
Figure4-10	Morlet Wavelet	65
Figure4-11	The stretched Haar wavelet filter;(a)16 points, (b)32 points, (c) 64 points, and (d) 128 points	66
Figure4-12	(a) The 4 points basic Db4 wavelet, (b) the Db4 after stretched to 764 points	67
Figure4-13	(a) The Db20 wavelet look likes the chirp signal, (b) The Haar wavelet match the binary signal	68

List of Figures (cont....)

Figure No.	Title	Page No.
Figure 5-1	The biological neuron	76
Figure 5-2	Artificial neuron model	79
Figure 5-3	Threshold activation function	80
Figure 5-4	Piecewise-linear activation function	81
Figure 5-5	Sigmoid activation function	82
Figure 5-6	Single layer network	83
Figure 5-7	Multi-layer network	84
Figure 5-8	A single-loop feedback network	84
Figure 5-9	Recurrent network with self-feedback	85
Figure5-10	Error-Correction learning	90
Figure5-11	Competitive Learning	93
Figure 6-1	The proposed speaker recognition system	96
Figure 6-2	The speech signal for the first speaker before and after normalization	99
Figure 6-3	The db20 wavelet (a)symmetric nature, (b) their four filters	101
Figure 6-4	Energy distribution of sentence 1 for the two different speakers	102
Figure 6-5	The FFBPNN structure	103
Figure 6-6	The confusion matrix	104
Figure 7-1	Speaker recognition circuit	111
Figure 7-2	Block diagram of the speaker recognition circuit	112
Figure 7-3	ATmega328P pin configuration	114
Figure 7-4	EasyVR Commander	115

List of Tables

Table No.	Title	Page No.
Table 4-1	Fourier Transforms and the corresponding signal	36
Table 6-1	Description of the FFBPNN	104
Table 6-2	The target matrix	105
Table 6-3	Recognition rate results at the Training stage	106
Table 6-4	some of the weights values from the second hidden layer to the output layer	107
Table 6-5	recognition rate results at the Testing stage	109
Table 6-6	Comparison with MFCC, Zak in terms recognition rate	109
Table 7-1	List of pins assignment of the speaker recognition circuit	113

List of Abbreviations

ADALINE	ADaptive LInear Neuron
ANN	Artificial Neural Network
ANNs	Artificial Neural Networks
ART	Adaptive Resonance Theory
ASRS	Automated Speaker Verification System
CWT	Continuous Wavelet Transform
Db	Daubechies
db20	Daubechies 20-tap
DFT	Discrete Fourier Transform
DTFT	Discrete Time Fourier Transform
DWT	Discrete Wavelet Transforms
ELSDSR	English Language Speech Database for Speaker Recognition
FFBPNN	Feed-Forward Back-Propagation Neural Network
FFT	Fast Fourier Transform
IDWT	Inverse Discrete Wavelet Transform
IMM	Informatics and Mathematical Modeling
LMS	Least Mean Square
MADALINE	Multiple- ADaptive LInear NEuron
MFCCs	Mel Frequency Cepstral Coefficients
MRA	Multi-Resolution Analysis
RBF	Radial Basis Function
SIS	Speaker Identification System
SOM	Self-Organization Map
SRS	Speaker Recognition System
STFT	Short Time Fourier Transform
SVM	Support Vector Machines
SVS	Speaker Verification System
WPT	Wavelet Packet Transform
