



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

١- Prof. Dr. Ashraf shamseldin Yahia

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

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

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

٣- Dr. El-Sayed A. El-Dahshan

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

٢٠١٦

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 90.7% 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	IV
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

2,0,1,1 <i>The Ear</i>	24
2,0,2 <i>The brain</i>	20
2,6 Generic Speaker Recognition System	26
2,6,1 <i>Pre-processing</i>	27
2,6,2 <i>Feature Extraction</i>	27
2,6,3 <i>Modeling</i>	27
2,6,4 <i>Matching</i>	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 <i>Feature Extraction Techniques</i>	31
2,10,2 <i>Feature Matching Techniques</i>	34
Chapter 3 Signal Pre-processing	30
3.1 Introduction	30
3.2 Signal Pre-processing	30
3.2.1 <i>Digitalization</i>	36
3.2.2 <i>Truncation</i>	37
3.2.3 <i>Windowing</i>	38
3,2,3,1 <i>Convolution</i>	38
3,2,3,2 <i>Correlation</i>	39
3.2.4 <i>Energy Normalization</i>	41
Chapter 4 Wavelet Transform	43
4.1 Introduction	43

ξ.ϒ Historical Development	ξ ξ
ξ.ϓ Fourier Analysis	ξ Ϟ
ξ.ϓ.ϑ Fourier Transform.....	ξ Ϡ
ξ.ϓ.ϒ Fourier Series	ξ ϡ
ξ.ϓ.ϓ Discrete Fourier Transform.....	ξ ϡ
ξ.ϓ.ϔ Short Time Fourier Transform.....	ο ϑ
ξ.ϔ What are Wavelets?	ο ϒ
ξ.ο Wavelet Transform	ο ο
ξ.ο.ϑ The Continuous Wavelet Transform.....	ο Ϟ
ξ.ο.ϒ The Discrete Wavelet Transform.....	ο Ϡ
ξ,ο,ϒ,ϑ <i>Multi-Resolution Analysis</i>	Ϟ Ϟ
ξ.ο.ϓ Wavelet Packet Transform.....	Ϟ ϑ
ξ.Ϟ Similarities to Fourier Transform	Ϟ ϓ
ξ.ϟ Some of Most Popular Wavelet Types	Ϟ ϓ
ξ.ϟ.ϑ Daubechies Wavelet.....	Ϟ ξ
ξ.ϟ.ϒ Mexican Hat Wavelet.....	Ϟ ξ
ξ.ϟ.ϓ Morlet Wavelet	Ϟ ο
ξ.Ϡ Examples of Wavelet Filters	Ϟ ο
ξ.ϡ Choosing the Best Wavelet	Ϟ ϕ
ξ.ϡ Ϟ Wavelet Applications	Ϟ Ϡ
Chapter ο Artificial Neural Network.....	ϕ Ϟ
ο.ϑ Introduction	ϕ Ϟ
ο.ϒ What are Artificial Neural Networks (ANNs)	ϕ Ϟ
ο.ϓ Characteristics of Artificial Neural Networks.....	ϕ ϒ

○.ξ History and Development of Artificial Neural Networks	۷۳
○.ο Biological Neuron	۷۵
○.Ϛ The Artificial Neuron	۷۷
○.Ϛ.۱ The Neuron Model.....	۷۷
○.Ϛ.۲ Some Activation Functions.....	۷۹
○.ϣ Artificial Neural Networks Architecture.....	۸۲
○.ϣ.۱ Feed-forward Networks.....	۸۲
○,ϣ,۱,۱ <i>Single Layer Feed-forward Network</i>	۸۲
○,ϣ,۱,۲ <i>Multi-layer Network</i>	۸۳
○.ϣ.۲ Recurrent Networks.....	۸۴
○.λ Learning Process	۸۵
○.λ.۱ Learning Paradigms (models).....	۸۶
○,λ,۱,۱ <i>Supervised</i>	۸۶
○,λ,۱,۲ <i>Non supervised</i>	۸۷
○.λ.۲ Some Learning Algorithms	۸۷
○,λ,۲,۱ <i>Back-propagation Algorithm</i>	۸۷
○.λ.۳ Learning Rules	۸۹
○,λ,۳,۱ <i>Error-Correction Rule</i>	۸۹
○,λ,۳,۲ <i>Hebbian Rule</i>	۹۰
○,λ,۳,۳ <i>Competitive Rule</i>	۹۲
○.ϩ Artificial Neural Networks Applications	۹۳
Chapter ۶ Developed System and Results.....	۹۵
۶.۱ Introduction.....	۹۵
۶.۲ The Proposed Speaker Recognition System.....	۹۵

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 V Hardware implementation	110
V.1 Introduction.....	110
V.2 Introducing the Speech Recognition Module	111
V.3 EasyVR ^{2,0} Specifications.....	112
V.4 System Setup.....	113
V.4.1 Arduino Board Specifications	114
V.5 System Operation	115
V.6 Results	116
Chapter ^ 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
Figure 4-10	Morlet Wavelet	65
Figure 4-11	The stretched Haar wavelet filter;(a) 16 points, (b) 32 points, (c) 64 points, and (d) 128 points	66
Figure 4-12	(a) The 4 points basic Db^4 wavelet, (b) the Db^4 after stretched to 16 points	67
Figure 4-13	(a) The Db^{10} 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
Figure 5-10	Error-Correction learning	90
Figure 5-11	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 db ₂₀ 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	ATmega ³² P 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 LLinear Neuron
ANN	Artificial Neural Network
ANNs	Artificial Neural Networks
ART	Adaptive Resonance Theory
ASRS	Automated Speaker Verification System
CWT	Continuous Wavelet Transform
Db	Daubechies
dbψ	Daubechies ψ -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 LLinear 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