



**Computer Science Department
Faculty of Computer & Information Sciences
Ain Shams University**

Biometric Intelligent System based on Heart Signals

Thesis submitted as a partial fulfillment of the requirements for the degree of
Master of Science in Computer and Information Sciences.

By

Mahmoud Mohamed Bassiouni EL-Sayed

Demonstrator at Computer Science Department,
Faculty of Computer and Information Sciences,
Egyptian E-learning University

Under Supervision of

Prof. Dr. Abdel-Badeeh Mohamed Salem

Professor of Computer Science,
Faculty of Computer and Information Sciences,
Ain Shams University

Prof. Dr. El-Sayed A. El-Dahshan

Professor of Scientific Computing,
Faculty of Sciences-Physics Department,
Ain Shams University

Dr. Wael Hamdy Khalifa

Computer Science Department,
Faculty of Computer and Information Sciences,
Ain Shams University

December 2016

Acknowledgment

I would like express my gratitude and a special thanks to my supervisor Prof. Dr. Abdel-Badeeh Mohamed Salem for his guidance, advice and constant support throughout my thesis. I would like to thank him for being my advisor. I want to express my gratitude to Prof. Dr. El-Sayed A. El-dahshan for his full support, encouragement, knowledge, teachings and his expected help. I want to thank Dr. Wael khalefa for his support in my thesis. I want to thank my friends and my workmates for all the thoughtful and motivating discussions. I am especially grateful to my parents for their love and support and would like to thank my parents for raising me in a way to believe that I can achieve anything in life with hard work and patience.

Abstract

Recent reported research proved that Electrocardiogram (ECG) and Phonocardiogram (PCG) can be used as a biometric. In this work, we built and developed two different biometric identification systems one for (ECG) and the other for (PCG).

In the first system, we presented an ECG divided mainly into four stages namely, data acquisition, preprocessing, feature extraction and classification. First stage, data acquisition stage, data sets were collected from two different databases, ECG-ID and MIT-BIH Arrhythmia database. Second stage noise reduction of ECG signals using wavelet transform and series of filters used for de-noising. Third stage obtains the features using three different techniques a non-fiducial, fiducial and a fusion approach between them. In the last stage, the classification stage, three classifiers have been developed to classify subjects. The first classifier is based on Artificial Neural Network (ANN), the second classifier is based on Euclidean distance (ED) and the last classifier is sequential minimal optimization (SMO) algorithm for training a support vector machine (SVM) using polynomial kernel classifier. Classification accuracy of 95% for ANN, 99 % for ED and 99% for SVM on the ECG-ID database, while 100% for ANN, ED, SVM on MIT-BIH database.

In the second system we presented a machine learning approach based on feature level fusion for person identification using phonocardiogram (PCG). The proposed approach consists of five stages; starting with data acquisition, pre-processing, segmentation, feature extraction, and classification. Firstly data set were collected from the HSCT-11 database working on 60 subjects. Secondly process is concerned with noise reduction by removing noise from PCG signal

using wavelets. Thirdly, segmentation process was done by applying Shannon energy envelope on the filtered signal to detect the positions of S1 and S2. Then a new combination of features is investigated for robust biometric PCG identification. Mel Frequency Cepstral Coefficients (MFCC) is efficient for PCG identification in clean speech while Wavelet features are robust for noisy environments. Therefore, combining both features together is better than taking each one individually the fusion is done using canonical correlation analysis (CCA). Finally, artificial neural network (ANN) and sequential minimal optimization (SMO) algorithm for training a support vector machine (SVM) using polynomial kernel classifier has been applied to classify subjects. The result shows a classification ratio 98.33%.

The results obtained from those two systems (ECG) and (PCG) were encouraging to show how robust our machine learning techniques used are. A comparison is already made with the previous work to ensure the efficiency of our methods.

List of Publications

- 1- BASSIOUNI M., KHALEFA W., EL-DAHSHAN E.S.A., SALEM A.B.M., A Machine Learning Technique for Person Identification using ECG Signals, International journal of Applied physics, Vol 1 2016 (pp. 37-41).
- 2- Bassiouni M., Khalefa W., El-Dahshan E.S.A. and Salem A.B.M., A study on the Intelligent Techniques of the ECG-based Biometric Systems. *RECENT ADVANCES in ELECTRICAL ENGINEERING*, 2015, (pp.26-31).
- 3- Bassiouni, M., Khalifa, W., El Dahshan, E.S.A. and Salam, A.B.M., A study on PCG as a biometric approach. In *2015 IEEE Seventh International Conference on Intelligent Computing and Information Systems (ICICIS)*, 2015, December, (pp. 161-166).
- 4- Bassiouni, M., Khalifa, W., El Dahshan, E.S.A. and Salam, A.B.M., An Intelligent Approach for Person Identification Using Phonocardiogram Signals, International Journal of Applications of Fuzzy Sets and Artificial Intelligence (ISSN 2241-1240), Vol. 6, 2016, (pp.103 -117).

Table of Contents

List of Publications	5
Table of Contents	6
List of Figures	8
List of Tables	10
List of Algorithms.....	11
List of Abbreviations	12
Chapter1.Introduction	14
1.1 Overview	14
1.2 Motivations for ECG and PCG	14
1.3 Objectives.....	15
1.4 Biometrics	16
1.5 Contributions.....	26
1.6 Thesis Organization	27
Chapter 2.Biomedical Background.....	29
2.1 Introduction to ECG Signal.....	29
2.2 ECG as a biometric.....	33
2.3 Introduction to PCG Signal	35
2.4 PCG as a biometric	38
Chapter 3.A study on the Intelligent Techniques of Electrocardiogram -based Biometric Systems	41
3.1 Introduction	41
3.2 General Methodology of ECG biometric.....	41
3.3 Background of the Techniques used in ECG	45
3.4 Analysis of the Techniques used in ECG Recognition.....	50
3.5 Discussion	55
3.6 Summary	56
Chapter 4.A study on phonocardiogram as a biometric	58
4.1 Introduction	58
4.2 General Methodology of PCG biometric.....	58
4.3 Background of the Techniques used in PCG	62
4.4 Analysis of the Techniques used in PCG Recognition	64
4.5 Discussion	68
4.6 Summary	69
Chapter 5.Machine Learning Hybrid approaches for Electrocardiogram Human Identification	71

5.1	Introduction	71
5.2	Proposed Methodology	74
5.2.1	Data Acquisition Phase	76
5.2.2	Preprocessing Phase	76
5.2.3	Feature Extraction Phase	78
5.2.4	Classification Phase	91
5.3	Experiments and Results	95
Chapter 6.	Human Identification based on the heart sounds (PCG) signals	102
6.1	Medical Aspects	102
6.2	Related Work	104
6.3	Proposed PCG Identification System	105
6.3.1	Database	107
6.3.2	Preprocessing	107
6.3.3	Segmentation	109
6.3.4	Feature Extraction	111
6.3.5	Classification	118
6.4	Results and Discussion	122
6.5	Summary	125
Chapter 7.	Conclusion and Future Work	127
7.1	Conclusion	127
7.2	Future Work	127
	References	128
	Arabic Abstract	138

List of Figures

Figure Number	Figure Title	Figure Page
1.1	Biometric Traits	15
1.2	Physical and Behavioral Biometrics	15
1.3	Operation methodology of Biometric systems	16
1.4	Artificial Noise used in Oder	20
2.1	Components of ECG signal	26
2.2	Electrode position for recording	27
2.3	Electrical event of the cardiac cycle	30
2.4	Heart Sounds S1,S2,S3 and S4	33
2.5	Four Auscultation sites	33
2.6	Actions of valves and chambers when ventricles are in diastole and systole	35
3.1	Ideal ECG signal containing P,Q,R,S,T,U peaks	38
3.2	General ECG Biometric Recognition System	39
3.3	Devices to capture the ECG Signal	40
3.4	8 Distances extracted from ECG	42
3.5	19 features extracted from ECG	43
3.6	21 features extracted from ECG	44
3.7	Different Wavelet Families	45
4.1	Normal PCG Signal	56
4.2	Methodology of PCG Biometric Recognition System Stages	57
5.1	Electrical Events of the cardiac cycle and their corresponding graph on the ECG	69
5.2	Proposed Methodology	72
5.3	ECG ID Person 1 Signal 1 Original Signal	74
5.4	ECG ID Person 1 Signal 1 Filtered Signal	74
5.5	MIT-BIH Person Subject 101 Original Signal	75
5.6	MIT-BIH Person Subject 101 Filtered Signal	75
5.7	(a) shows ECG-ID Subject 1 the preprocessed signal (b) Show the normalized autocorrelation sequence (c) 400 AC coefficients from the maximum peak in AC (d) 30 samples produced after applying DCT of the 400 AC coefficients	77
5.8	(a) shows MIT-BIH Subject 101 the preprocessed signal (b) Show the normalized autocorrelation sequence (c) 400 AC coefficients from the maximum peak in AC (d) 30 samples produced after applying DCT of the 400 AC coefficients	78
5.9	A vector of zeros and ones where ones represent QRS	78

5.10	(a) Shows ECG-ID Subject 1 the preprocessed signal detected peaks	79
	(b) Shows a zoomed ECG-ID Subject 1 the preprocessed signal detected peaks	
5.11	(a) Shows MIT-BIH Subject 101 preprocessed signal detected Peaks	79
	(b) Shows a zoomed MIT-BIH Subject 101preprocessed signal detected	
5.12	Variants of cardiac cycle information fragments	80
5.13	Test set ECG identification results for considered informative fragments	81
5.14	P-QRS-T Fragments Computed from ECD-ID Subject1	82
5.15	P-QRS-T Fragments Computed from MIT-BIH Subject 101	83
5.16	(a) Mean P-QRS-T Fragments Computed from ECD-ID Subject 1 (b) Mean P-QRS-T Fragments Computed from MIT-BIH Subject 101	83
5.17	(a) Mean P-QRS-T Fragments Computed from ECD-ID Subject 1 (b) Mean P-QRS-T Fragments Computed from MIT-BIH Subject 101	85
5.18	38 fiducial features for ECG signal	86
5.19	First Decomposition of DWT for ECG Signal	88
5.20	Accuracy performance of each classifier on each approach on ECG ID	93
5.21	Accuracy performance of each classifier on each approach on MIT-BIH Arrhythmia	94
6.1	Action of valves and chambers when the ventricles are in diastole (A) and systole (B)	100
6.2	Waveform of PCG Signal	101
6.3	Human Identification based on PCG Signals	103
6.4	a) Original PCG Signal, b) PCG Signal after adding Gaussian Noise, c) De-noised	105
6.5	a) Shannon Energy of PCG Signal (b) Normalized Shannon Energy of PCG	107
6.6	Steps of the MFCC feature extraction	109
6.7	Decomposition of PCG Signal using discrete wavelet transform	113
6.8	General Artificial Neural Network Structure	117
6.9	Accuracy of PCG signals for the two feature extraction algorithms MFCC+PCA and MSDWT with SVM and ANN	120
6.10	Accuracy of PCG signals fusion using Concatenation, CCA Serial and CCA Parallel with SVM and ANN	121

List of Tables

Table Number	Table Name	Table Page
1.1	Comparison of various biometric performance	25
3.1	Extracted attributes. The feature list labels are the normalized distance between the two fiducials. For example RP' is the unsigned distance between the end of the P-wave and the R peak	44
3.2	19 features extracted from ECG	45
3.3	21 features extracted from ECG	45
3.4	Machine learning Techniques in ECG	51
4.1	Machine learning Techniques in PCG	65
5.1	Labels of the 38 fiducial features of our modified super set	88
5.2	Parameters used in ANN for ECG Classifications	93
5.3	TP and FP for each of the three Classifiers and the three feature extraction approaches on the ECG-ID database	94
5.4	Classification Results for ECG-ID with the three approaches in feature extraction and classification	94
5.5	TP and FP for each of the three Classifiers and the three feature extraction approaches on the MIT-BIH database	94
5.6	Classification Results for MIT-BIH with the three approaches in feature extraction and classification	95
5.7	Classification performance comparisons of the proposed Scheme with some existing schemes using MIT-BIH Database	96
5.8	Classification performance comparisons of the proposed Scheme with some existing schemes using ECG-ID Database	97
6.1	Parameters used in ANN for PCG classification	119
6.2	Parameters used in SVM for PCG classification	121
6.3	Correctly Classified Subjects using MFCC, MSDWT	121
6.4	Accuracy performance of the system for MFCC, MSDWT	122
6.5	Correctly Classified from the fusion of the system based on MFCC, MSDWT	122
6.6	Accuracy performance of the fusion based on MFCC, MSDWT	122
6.7	Comparison between the proposed system and the previously implemented PCG biometric authentication systems selected from HSCT-11 database	124

List of Algorithms

Algorithm Number	Algorithm Title	Algorithm Page
5.1	Steps for Detecting most discernment P-QRS-T for ECG Identification	84
6.1	PCG Signal De-noising	106
6.2	PCG Signal Segmentation using Shannon Energy	108
6.3	MFCC Feature Extraction Algorithm	111
6.4	MSDWT Feature Extraction Algorithm	113
6.5	Fusion between MFCC and MSDWT	117

List of Abbreviations

ANN	Artificial Neural Network	PRD	Percent Residual Difference
AC	Auto Correlation	PSA	Power Spectral Analysis
BP	Back Propagation	PTB	Physikalisch Technische Bundesanstalt
BFCC	Bark Frequency Cepstral Coefficient	QNN	Quantum Neural Network
CRR	Correct Recognition Rate	QP	Quadratic Programming
CCA	Canonical Correlation Analysis	PCA	Principle Component Analysis
CCORR	Correlation Coefficient Recognition Rate	RBFNN	Radial Bias Function Neural Network
DNA	Deoxyribo Nucleic acid	RS	Rough Set
DTW	Dynamic Time Wrapping	SMO	Sequential Minimal Optimization
DCT	Discrete Cosine Transform	SVM	Support Vector Machine
DWT	Discrete Wavelet Transform	SACA	Statistical and Coherence Analysis
DNN-BT	Deep Neural Network Breaking Tires	SOM	Self-Organization Map
DBNN	Decision Based Neural Network	STDFT	Short Time Discrete Fourier Transform
ECG	Electrocardiogram	TM	Template Matching
ED	Euclidean Distance	TC	Template Correlation
ECG-ID	Electrocardiogram – Identification	TPR	True Positive Rate
FFT	Fast Fourier Transform	WT	Wavelet Transform
FPR	False Positive Rate	WDIST	Wavelet Distance
GMM	Gaussian Mixture Models	WPT	Wavelet Packet Transform
HHT	Huang Hilbert Transform	VQ	Vector Quantization
HMM	Hidden Markov Models		
HSCT-11	Heart Sound Catania 2011		
ISOM	Increment Self Organization Map		
KLDA	Kernel Linear Discernment Analysis		
KNN	K-Nearest Neighbor		
LBFC	Linear Band Frequency Cepstra		
LPCC	Linear Predication Cepstral Coefficient		
LDA	Linear Discernment Analysis		
MIT-BIH	Massachusetts Institute of Technology		
MSE	Mean Square Error		
MSDWT	Multi Scale Discrete Wavelet Transform		
MLP	Multi-Layer Perceptron		
MVC	Majority Vote Classifier		
MFCC	Mel Frequency Cepstral Coefficient		
NN	Neural Network		
PCG	Phonocardiogram		
PBNN	Probability Based Neural Network		
PNN	Probabilistic Discernment Analysis		
PSA	Power Spectral Analysis		

Chapter 1

Introduction

Introduction

1.1 Overview

Due to identity fraud, loss of privacy, friendly thief, terrorism the identification security is considered one of the important solutions to these problems. Although computers help human and provide an important services to him but also they does not deliver enough security solutions to identify the individual. Systems based on computer for identification are defined as biometric systems. Those systems are used a lot now a days and in a lot of applications [1].

1.2 Motivations for ECG and PCG

Some people are still reluctant to engage in ecommerce or conduct other network transactions having misgivings about well-founded systems that will protect their privacy and prevent their identity from being stolen or misused. The loss of personal privacy, fraudulent funds transfers, outright theft, and abuse of identity in network transactions leads to identity fraud and identity theft. Identity theft is defined as the personal information is accessed by third parties without obvious authorization from the owner. Identity fraud takes place when a criminal gets an illegally-obtained personal information and uses it for his financial gain. Those personal information may comprise social insurance number, bank or credit card account numbers, passwords, telephone calling card number, birth date, name, address and so on.

Another type of identity theft is known as online identity theft such as phishing, hacking and spyware. Also friendly theft that occurs within friends, family or in-home employees who obtain private data for their personal gain. Due to all those theft problems researchers were thinking and researching for a lot of solutions. This motivated researchers to seek for a dependable and exact substitute solution for identity identification over the traditional password/ ID card based systems and any traditional techniques used [2].

Biometrics, which are physiological or behavioral characteristics extracted from human subjects, have emerged to be a new set of technologies that promise an effective solution for this problem. So they cannot be lost, stolen, forged, or subject to failure. In the past few periods,

biometrics has been lengthily used for law execution such as criminal investigations, fatherhood determination, and forensics [3].

On one hand the traditional techniques using username and password, also using keys and identification cards can be stolen and lost and also the traditional biometrics such as face, Signature, DNA, Speech, Fingerprint, face and iris they all can be falsified, forged, fooled and faked. On the other hand PCG and ECG biometric are expected to address some deficiencies thanks to their universality, distinctiveness, permanence, small storage requirements, speed, and their stability over a sufficiently long period of time as well as their uniqueness for individuals. Those signals are difficult to disguise, forged, falsified or faked.

1.3 Objectives

- The Main Aims of the thesis is the analysis of two relatively new medical biometric attributes, the electro-cardiogram (ECG), and the heart sound (PCG),

The primary objectives of the current research are to address the following issues:

- **Data collection:** Searching for a dataset recently in ECG and PCG for patients of in different ages, currently searching for a combined database for both ECG and PCG for the same patient.
- **Design of a robust Identification system based on ECG signal:** To date, two major approaches have been proposed for ECG recognition, namely fiducially-based and non-fiducially-based methods, that achieve good performance in an almost ideal scenario in terms of noise and heart rate.
- **Design of a robust Identification system based on PCG signal:** Human recognition based on the PCG signals is a new area only lately. Having an acoustic nature, the use of the PCG signal in recognition context is very challenging due to wave characteristic in the presence of noise components.