A BIOMETRIC AUTHENTICATION TECHNIQUE USING SPREAD SPECTRUM AUDIO WATERMARKING

by

Yekta Said CAN

B.S., Computer Engineering, Boğaziçi University, 2012

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Computer Engineering

Boğaziçi University

2014

A BIOMETRIC AUTHENTICATION TECHNIQUE USING SPREAD SPECTRUM AUDIO WATERMARKING

APPROVED BY:

Prof. Fatih ALAGÖZ .....
(Thesis Supervisor)

Prof. Fikret GÜRGEN .....

Prof. Emin ANARIM .....

DATE OF APPROVAL: 28.01.2014

ACKNOWLEDGEMENTS

I would like to thank my thesis supervisor Prof. Fatih Alagöz. His encouragement and rigorous , continuous guidance made the research easier. I have been really lucky to know him and work with him.

My deepest gratitude goes to my family, my parents Muhiddin , Müşerref CAN and my little sister Asiye Sevde CAN for their unflagging love and support throughout my life.

I also would like to express my deep and sincere gratitude to all my friends Mert,Ömer,Hakan,Turgut,Burak. I want to thank these nice people for the pleasant time I had with them.

ABSTRACT

A BIOMETRIC AUTHENTICATION TECHNIQUE USING SPREAD SPECTRUM AUDIO WATERMARKING

Watermarking has become important in the last decade because of the copyright protection applications. Embedding information into an audio file is more difficult as compared to images, because human auditory system is more sensitive than human visual system. Therefore, the proposed watermarking algorithms for digital audio have been less than those for digital image and video. This thesis presents a biometric authentication scheme based on spread spectrum watermarking technique. We add a biometric authentication system to the Sipdroid open source VoIP program. Firstly, senders must register to the system with their unique biometric features. T.C Identity number, keystroke dynamics and voice are used as biometric features. After registration, these biometric features are used as watermarked material. Before embedding, the watermark is spread with the Direct Sequence Spread Spectrum (DSSS) technique. While talking, this watermark material is embedded to speech and sent to receiver using Frequency Hopping Spread Spectrum(FHSS) technique. The watermarked biometric data is constructed in the receiver's phone after conversation is finished. This method does not need the original audio carrier signal when extracting watermark because it is using the blind extraction. The experimental results demonstrate that the embedding technique is not only less audible but also more robust against the common signal processing attacks like low-pass filter, adding white Gaussian noise, shearing, and compression. In order for receiver to be able to login to the system, biometric features of the user should match with the watermarked biometric data.

ÖZET

GERÇEK ZAMANLI GENİŞ SPEKTRUMLU DİJİTAL SES DAMGALAMA KULLANARAK BİYOMETRİK KİMLİK DOĞRULAMA

Telif hakkı uygulamalarının artmasından dolayı son on yılda damgalama sistemleri önem kazandı. İnsan duyma sistemi görme sistemine göre daha hassas olduğundan sese bilgi gömmek resimlere kıyasla daha zordur. Bundan dolayı dijital ses için damgalama algoritmaları resimler için olanlardan daha azdır. Bu tezde geniş spektrum dijital ses damgalama sistemi kullanılarak biyometrik kimlik doğrulama sistemi sunuldu. Açık kaynak IP üzerinden ses gönderme programı olan Sipdroid programı Android işletim sistemi üzerinde kullanıldı. Biyometrik kimlik doğrulama sistemi Sipdroid programına eklendi. Öncelikle, konuşmada gönderen kısım sisteme eşsiz biyometrik özellikleri ile kayıt olmalıdır.T. C. kimlik numarası, tuş basma dinamikleri ve ses biyometrik özellik olarak seçildi. Kayıttan sonra, kaydedilen biyometrik özellikler sese damgalanan materyal olarak kullanıldı.Gömme işleminden önce imge Doğrudan Serili Geniş Spektrum (DSSS) sistemi kullanılarak genişletildi. Konuşma sırasında imge konuşmaya Frekans Sekmeli Geniş Spektrum (FHSS) sistemi kullanılarak gömüldü ve alıcıya gönderildi. Damgalanan ibiyometrik veriler konuşma sonrası alıcının telefonunda yeniden oluşturuldu. Uygulanan metod orijinal ses sinyaline ihtiyaç duymaz. Deneysel sonuçlar gömme sisteminin hem daha az duyulabilir hem de gauss gürültüsü ekleme, düşük geçiş filtresi, kesme, sıkıştırma gibi yaygın işaret işleme ataklarına karşı daha sağlam olduğunu gösterdi. Alıcı kısmında sisteme giriş yapılabilmesi için, kullanıcının biyometrik özelliklerinin damgalanan biyometrik veri ile eşleşmesi gerekir.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii
ABSTRACT iv
ÖZET v
LIST OF FIGURES ix
LIST OF TABLES xi
LIST OF SYMBOLS xii
LIST OF ACRONYMS/ABBREVIATIONS xiii
1. INTRODUCTION 1
1.1. Background 1
1.2. Watermarking 2
1.3. Applications of Watermarking 3
1.3.1. Owner Identification 3
1.3.2. Broadcast Monitoring 3
1.3.3. Content Authentication 4
1.3.4. Copy Control 4
1.3.5. Information carrier 4
1.4. Outline of the Thesis 5
2. AUDIO WATERMARKING TECHNIQUES 7
2.1. Requirements of the Efficient Watermark Technique 7
2.1.1. Imperceptibility 7
2.1.2. Robustness 7
2.1.3. Capacity 8
2.2. Attacks on Audio Signals 8
2.2.1. Low Pass Filter 8
2.2.2. Shearing 9
2.2.3. Lossy Compression 9
2.2.4. Additive White Gaussian Noise 9
2.3. Audio Watermarking Techniques - An Overview 9
2.3.1. LSB Coding 11

2.3.2. Spread Spectrum Technique 12
2.3.2.1. Direct Sequence Spread Spectrum 12
2.3.2.2. Frequency Hopping Spread Spectrum 12
2.3.3. Transformation Techniques 13
2.3.3.1. Discrete Wavelet Transform 13
2.3.3.2. Haar Wavelet 15
3. BIOMETRIC AUTHENTICATION TECHNIQUES 17
3.1. Common Biometric Characteristics 17
3.2. Biometric Systems 18
3.3. Different Types of Biometric Techniques 18
3.3.1. Finger Print Recognition 18
3.3.2. Face Recognition 19
3.3.3. Iris Recognition 20
3.3.4. Hand Geometry 22
3.3.5. Voice Biometrics 22
3.3.6. Keystroke Dynamics 24
3.4. Working of Biometrics 25
3.5. Biometric Performance Measures 25
3.6. Application of Biometric Techniques 26
4. PROPOSED TECHNIQUE FOR WATERMARKING 27
4.1. Infrastructure 27
4.2. Time Domain FHSS Method [37] 28
4.2.1. Watermark Preprocessing 28
4.2.2. Watermark Embedding 29
4.2.3. Watermark Extracting 31
4.3. Wavelet Domain FHSS and DSSS Method[38] 32
4.3.1. Watermark Preprocessing 32
4.3.2. Watermark Embedding 33
4.3.3. Watermark Extraction 34
5. COLLECTING PERSONAL BIOMETRIC DATA 36
5.1. Registration 37

5.1.1. Unique TC Identity Number 37
5.1.2. Collecting Keystroke Dynamics Data 38
5.1.3. Biometric Feature Extraction From Voice 42
5.1.3.1. Zero Crossing Rate Calculation 42
5.1.3.2. Standard Deviation Calculation 43
5.1.3.3. Average Magnitude Calculation 43
5.1.3.4. Acceptance Interval Calculation 43
5.2. Login 45
5.2.1. TC Identity Number Check 45
5.2.2. Keystroke Dynamics Authentication 45
5.2.3. Voice Authentication 45
6. EXPERIMENTS AND RESULTS 48
6.1. Watermarking System Experiments 48
6.1.1. Inaudibility 49
6.1.2. Robustness 50
6.1.2.1. Compression Attack 50
6.1.2.2. AWGN Attack 50
6.1.2.3. Shearing and Low Pass Filter Attacks 51
6.2. Biometric Authentication System Experiments 52
6.2.1. FAR and FRR Rates 52
7. CONCLUSION 56
REFERENCES 58

LIST OF FIGURES

Figure 1.1. Digital Watermark Embedding. 2

Figure 1.2. Digital Watermark Extracting. 3

Figure 2.1. LSB Embedding [3]. 11

Figure 2.2. DSSS [19]. 12

Figure 2.3. FHSS [19]. 13

Figure 2.4. Multi-Level DWT [3]. 15

Figure 3.1. A Fingerprint Image. 19

Figure 3.2. Face Recognition Features [25]. 21

Figure 3.3. Iris recognition procedure [27]. 21

Figure 4.1. Desired Communication Channel. 28

Figure 4.2. Embedding with Hopping Sequence 23345. 29

Figure 5.1. You can either begin to create your biometric data by pressing the button register or try to login with your existing data. 36

Figure 5.2. After entering Identity Number press Confirm and then press Next. 37

Figure 5.3. Keystroke Dwell Time and Flight Time [17]. 39

Figure 5.4. After entering your password press next. After Wrong Attempts a Warning Appears. 41

Figure 5.5. Users record their voice sample five times. 42

Figure 5.6. User arrives at calling activity after registration and sends the biometric data. 44

Figure 5.7. The message indicating that password is right but behaviour is different. 46

Figure 5.8. The message indicating that voice of user is different. 47

Figure 6.1. Correlation Value versus Compression Ratio. 51

Figure 6.2. Normalized Correlation versus Gaussian Noise (dB). 52

Figure 6.3. Correlation Values against Shearing Attack Ratio (%). 53

Figure 6.4. Extracted Features From Four Users. 54

LIST OF TABLES

Table 6.1. SNR of Watermarked Audio. 49

Table 6.2. Correlation Values When LPF and Shearing Applied. 53

Table 6.3. False reject rate (FRR). 54

Table 6.4. False Acceptance Rate (FAR). 55

LIST OF SYMBOLS

D_t	Average of the dwell times
e_f	Difference of unfiltered dwell time from the average
e_d	Difference of unfiltered flight time from the average
E_{Temp}	Highest difference between the average and unfiltered dwell and flight times
E_n	watermarked host audio signal
$E^d_{Dynamic}$	Dynamic margin of error for dwell latency
$E^f_{Dynamic}$	Dynamic margin of error for flight latency
E_{User}	total margin of error for a user
$E_{Default}$	estimated default margin of error
e_n	watermark extracted from the attacked host signal
F_t	Average of the flight times
$H1(i)$	Detailed coefficient from the first level DWT Transform
I_n	original host audio signal
$L1(i)$	Approximate coefficient from the first level DWT Transform
L	Length of the signal
$ones[n]$	The number that 'OR'ed with the amplitude to embed 1 bit
$S(i)$	Time domain representation of the signal
S_x	x^{th} amplitude of the signal
Th^d_{Lower}	Dwell Time lower threshold
Th^f_{Lower}	Flight Time lower threshold
Th^d_{Upper}	Dwell Time upper threshold
Th^f_{Upper}	Flight Time upper threshold
w_n	original watermark signal
$zeros[n]$	The number that 'AND'ed with the amplitude to embed 0 bit
ρ	correlation factor
σ	standard deviation

LIST OF ACRONYMS/ABBREVIATIONS

AM	Amplitude Modulation
AWGN	Additive White Gaussian Noise
CWT	Continuous Wavelet Transform
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DSSS	Direct Sequence Spread Spectrum
DWT	Discrete Wavelet Transform
FAR	False Acceptance Rate
FHSS	Frequency Hopping Spread Spectrum
FRR	False Rejection Rate
GMM	Gaussian Mixture Model
ICA	Independent Component Analysis
IP	Internet Protocol
JFA	Joint Factor Analysis
LDA	Linear Discriminate Analysis
LFCC	Linear Frequency Cepstral Coefficients
LSB	Least Significant Bit
MFCC	Mel Frequency Cepstral Coefficients
PCA	Principal Component Analysis
PLDA	Probabilistic Linear Discriminant Analysis
PN	PseudoRandom Noise
SD	Standard Deviation
SNR	Signal to Noise Ratio
SVM	Support Vector Machine
ZCR	Zero Crossing Rate

1. INTRODUCTION

Recently with the improvement of digital multimedia and internet technologies, the distribution of audio, image and video data has increased enormously because the distribution has become instantaneous and cheap. This increase has raised the question of authorization protection and copyrights. Content owners have suffered from the piracy. Thus, content owners have been searching technologies that help them to protect their rights. The sudden interest in watermarking has mainly started for this reason [1]. Because the music industry is one of the most profitable industries in the market, audio watermarking gains more importance. Despite the fact that main reasons for the research in watermarking are copy prevention and copyright protection, there are also other applications that watermarking can be utilized. Broadcast monitoring, transaction tracking, authentication, copy control, and device control are the other industrial applications of watermarking [2].

1.1. Background

When a technology comes with great advantages, usually it has its own drawbacks. Globalization and internet make information sharing and research easier. However, they also facilitate the work of malicious users to attack and pirate the digital media [3]. Watermarking process was first employed on images and was called as Image Watermarking. Image Watermarking is most commonly applied; and challenges were created by hackers. After that, another digital embedding technique which uses source as audio is developed and called as Audio Watermarking. It has gained importance among developers because it is used to prevent copyrights of the music. Digital watermarking is a technique by which copyright information is embedded into the host signal in a way that the embedded information is imperceptible and robust against intentional and unintentional attacks [4].

1.2. Watermarking

Watermarking is a technique through which the secure information is carried without degrading the quality of the original signal [3]. This method comprises of dual parts:

(i) Embedding part,
(ii) Extraction part.

Embedded key is used as in case of a steganography. The key is utilized to improve security,by not allowing unsanctioned person to retrieve data. The embedded object is called watermark, the watermark embedding medium is called as the original signal and the modified object is called as embedded signal or watermarked data [5].

The embedding part, shown in Figure 1.1, takes watermark, original signal and watermarking key as the inputs and forms the watermarked data with these inputs. Whereas, the inputs for the extraction part are embedded signal and the key as shown in Figure 1.2 [5]. In non-blind techniques, the original signal is also needed for extraction.

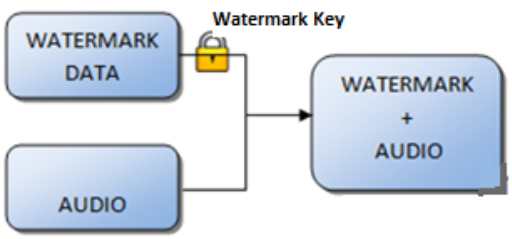

Figure 1.1. Digital Watermark Embedding.

Digital watermarking techniques can be mainly examined into two groups: blind and non-blind methods. A watermarking method that does not need original signal,

while extracting is called blind watermarking method [6]. Blind watermarking is more adventageous over nonblind techniques because watermarked signal and key are enough for extraction of embedded watermark.

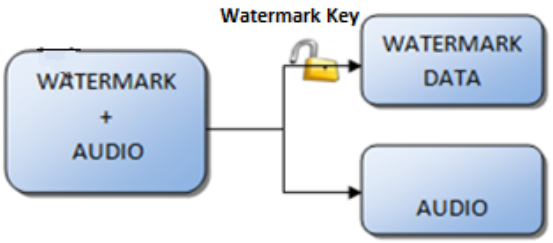

Figure 1.2. Digital Watermark Extracting.

1.3. Applications of Watermarking

1.3.1. Owner Identification

Owner identification was provided with textual copyright notices on image or audio. However there are problems with that approach. First problem is that the part which includes copyright notice can be removed or cropped. Second problem is that they can be aestehatically ugly and cover some part of the content [1]. Since watermark is imperceptible from the content and integrated to the content, it provides a better solution to this problem. Even if the content is altered, the owner can be still identified from the watermark.

1.3.2. Broadcast Monitoring

Over the last decade, the content that radio and television channels provide has increased enormously. It is more difficult for content owners to keep track of their media, whether their media are published or not and where they are published. Embedding watermark at production or broadcast time to audio or video allows content owners to know where the content is broadcast, who is broadcasting and for how long. The modifications are unseeable or inaudible. Watermark also prevents the

content from being edited. Moreover, it can be easily detected by hardware or software. Because it is part of the content, if a malicious user wants to remove, it will also ruin the original content [7].

1.3.3. Content Authentication

It is becoming easier to alter the digital contents. On most of the situations, we want to know whether the content we are dealing with is the original or altered. Along with the other ways, watermarking comes with a solution to this problem. The signature is embedded into the content. This is called the authentication mark [1]. It can be designed so that when someone changes the content it becomes invalid. So, modification becomes known.

1.3.4. Copy Control

In copy control applications,the aim is to prevent malicious users to make illegal copy of the copyrighted content. Because watermark exists in the content itself and they are at every presentation of it, it provides a better solution from just encryption [1]. If devices are equipped with a watermark detector, when the user tries to play the content that has 'do not copy' watermark, it checks whether it is original or copy. If this content is copy, the player will not play it.

1.3.5. Information carrier

The blind watermarking method might be applied in communication purposes. Watermark systems can be combined with communication systems as in our case. Conversation can be embedded with some content (images, text files). After that, receiving part can retrieve the watermark. This method can prevent malicious users from listening and retrive important information from conversation. Important information can be sent undetected via watermark.

1.4. Outline of the Thesis

The main idea of this thesis is to propose a real time digital spread spectrum audio watermarking technique. To improve security, a biometric authentication system is also added. Chapter 1 includes definition of watermarking, history and applications.

In Chapter 2, requirements of an efficient watermark technique are explained. The common attacks that can be applied on audio signals are also mentioned. Beginning with the simplest one, audio watermarking techniques are explained. First, least significant bit watermarking is illustrated. Then the spread spectrum watermarking which is applied in the thesis is explained. Finally, the most common transformation techniques, which are discrete cosine transformation watermarking and discrete wavelet transformation watermarking, are mentioned. The Haar Wavelet transformation is selected to explain among wavelets because this is the elementary technique and it is easy to understand.

Chapter 3 discusses the Sipdroid open source program, which we used to integrate the watermark system, in detail. Then the three stages of watermarking system are defined as watermark preprocessing, watermark embedding and watermark extracting respectively.

In Chapter 4, the components of biometric authentication systems are illustrated. The first one is T.C Identity number. The second component is the keystroke dynamics part. Which features we are extracted and how we use these features to login are shown. Finally, the voice authentication part is mentioned. The three features extracted from voice are explained which are zero crossing rate, standard deviation and average magnitude respectively.

Chapter 5 provides attacks which we applied on audio to demonstrate the robustness of the proposed technique. Firstly, SNR results that illustrate the inaudibility are shown. Then normalized correlation metric is utilized to evaluate the degradation of watermark after common signal processing attacks are applied. Three attacks are

selected which are adding AWGN noise, compression, shearing and low pass filtering respectively.

Chapter 6 concludes the thesis. Some future works that can be added to this research are also mentioned.

2. AUDIO WATERMARKING TECHNIQUES

2.1. Requirements of the Efficient Watermark Technique

Watermark systems should have certain properties. These properties depend on the application that watermark is used. In our application, three properties were taken into consideration: inaudibility, robustness and data payload [8].

2.1.1. Imperceptibility

One of the significant characteristics of a watermarking method is the inaudibility. Inaudibility (imperceptibility) is determined by the perceptual difference of the embedded watermark data from the original audio signal. Watermark should not deterioriate the original signal. The goal of inaudibility is that the quality of host signal is not corrupted much, so that the listener can not perceive the difference.

Imperceptibility requirement strictness changes from case to case. For example in the case of AM radio transmission, the quality of transmitted audio is low, so the watermark can become imperceptible after channel attenuations. However, in the case of HDTV or BluRay videos, since the quality of host video is too high, watermark should be as imperceptible as possible.

2.1.2. Robustness

Widespread signal processing operations like low pass filtering, lossy compression, shearing and AWGN can be implemented on watermarked host audio digital signal. Although effects of these attacks might not change the quality of the host signal much,embedded watermarking image is more impaired by these attacks. Robustness can be defined as the system's ability to detect the embedded watermark after common signal processing operations mentioned above [1].

The types of signal processing operations that a watermarking system should be robust against depend on the application. For example, watermarking systems that go through transmission like AM radio should be only robust against transmission process. However in the broadcasting digital video application, the system goes through analog to digital conversion and compression. So, the system should be robust against these operations. In addition, if there are malicious users trying to retreive the watermark, the system must be prepared for general signal processing attacks.

2.1.3. Capacity

The third property that a watermarking system should have is the data payload. Data payload is the amount of watermarked data at certain unit of time or image [1]. "Robustness and data rate are also important but these two cannot be achieved at the same time" [6]. So, in this thesis, particular importance is attached to robustness by making a little sacrifice from the data rate.

The definition of the capacity depends on the application. For an audio watermarking application the definition is the bits per second. For image watermarking it is bits per image. In a video watermarking it is either bits per second or bits per frame [1].

2.2. Attacks on Audio Signals

To measure the robustness of the watermarking method, several common signal processing attacks are applied [9].

2.2.1. Low Pass Filter

Filtering process is for removing some part of undesired piece of the signal. Filtering is very widespread, which can be utilized to improve or degrade some part of the signal. The basic low pass and high pass filters are utilized to apply these kinds of attacks. Here watermark signal is filtered with a Butterworth low-pass filter which is

second order and has a cutoff frequency of 0.5.

2.2.2. Shearing

Shearing is also another common signal processing attack. During transmissions, the part of transmitted audio might be lost due to channel conditions, jamming and so on. This attack is implemented, by assigning some part of watermarked audio to zero.

2.2.3. Lossy Compression

Lossy compression is another widely applied signal processing attack. It compresses data by discarding (losing) some part of it. The metric that determines how much compression affects data is the compression ratio. In other words, it means how much of the data is discarded during compression. Lossy compression with compression ratios 12.5, 25, 37.5, 50 per cent is applied. This attack is implemented by cutting the least significant bits then replacing them with random bits. For example, if compression ratio is 25 percent, then the last two bits of a byte are removed and then randomly replaced.

2.2.4. Additive White Gaussian Noise

The last common signal processing attack is adding an AWGN to the watermarked audio signal. During the transmission of the signal, there is almost always noise in the environment. So, in order to test the algorithm, noise should be added. Watermark algorithms should be robust against noise addition. The most common noise is additive white gaussian noise. The metric that determines how strong this attack is the signal to noise ratio. This attack is implemented on MATLAB by using the built-in function.

2.3. Audio Watermarking Techniques - An Overview

Watermarking techniques can be applied in time or transform domain. In time domain approach [10], information is embedded straight into the amplitudes of the

audio signal [6]. The second approach is the transform domain approach. Transform domain watermarking can be classified into three categories: Frequency domain, Discrete Cosine Transform (DCT) domain, Discrete Wavelet Transform (DWT) domain. In transform domain approaches [11, 12], the first step is the transformation of host signal into the specified domain and after that embedding information is carried out into the transformed media.

Spread spectrum method is one of the most important techniques for hiding data into an audio signal. Spreading nature of this technique makes the detection of embedded information almost impossible for a malicious user. There are two implementations of this technique: either using direct sequence spread spectrum approach or frequency hopping spread spectrum. In the direct sequence spread spectrum approach, Exclusive-OR operation is used for spreading the embedded information whereas in the frequency hopping spread spectrum technique, spreading is applied by using the hopping sequence [6].

There are a lot of spread spectrum techniques that are implemented. Here, we will describe a few of them. In [13], Darko Kirovski et al use an audio watermarking method where embedding is carried out in the time domain. Cepstral filtering method is used to embed the watermark bits into frames of audio. The carrier noise and large energy fluctuation issues can be reduced by this approach [6]. Extraction is done by utilizing normalized correlation technique after the embedding has finished. However, applied method does not achieve high robustness scores versus widespread signal processing attacks [6]. In the paper of Mark Sterling *et al.* [14], a Welch costas array is used in audio watermarking method.This array is a frequency hopped spread spectrum signal or a special type of steganographic signal [6]. Time domain is the embedding domain. A Match filter was used for extraction of the watermark. Discrete Fourier Transform based audio watermarking algorithm was used in Jared Vawter [15]. However, applied method does not achieve high robustness scores versus widespread signal processing attacks like MP3 compression [6]. Original host signal is necessary for extraction, so the scheme is non-blind. Discrete Fourier Transform based audio watermarking algorithm

was used in [4], Yiqin et al [12]. The later is the new version of the previous paper where DCT replaces DFT. Using DCT, robustness can be augmented,however the applied method does not achieve high SNR scores versus widespread signal processing attacks [6].

2.3.1. LSB Coding

This method is a widely applied method in signal processing. The decision of how many bits are being replaced affects the robustness. This technique is applied by substituting the least significant bits of the host signal with the bit sequence from embedded watermark. The bits are embedded into some representation values, such as pixels in image watermarking or amplitudes of sound in audio watermarking. The decoder then is able to extract the watermark if it knows the representation values used for embedding the individual bits[18].

The fundamental advantage of this scheme is its high watermark capacity, on the contrary; its main drawback is low robustness, because of random alteration of the LSBs demolish the embedded watermark [18]. It is hardly possible that a LSB encoded watermark will endure from compression and decompression operations. The characteristics of the LSB methods limit their applicability and a digital environment is required and common signal processing operations on the host signal can not be applied. Otherwise, operations corrupt the watermark.

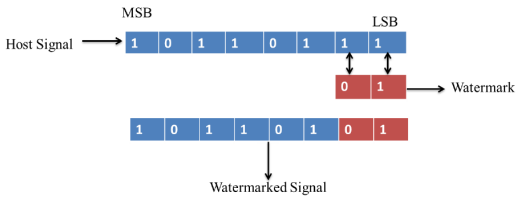

Figure 2.1. LSB Embedding [3].

2.3.2. Spread Spectrum Technique

Spread Spectrum technique is originally designed to prevent communication from jamming attacks. However, it has been utilized in watermarking area also. There are two types of spread spectrum technique.

2.3.2.1. Direct Sequence Spread Spectrum. The direct sequence spread spectrum (nSSS) method widens the bandwidth of the original signal. In this technique, an n bit spreading code is generated. Then, every bit in the watermark data is spreaded with this code. Every '1' is replaced with the negation of the nbit chipping sequence whereas each '0' is replaced with the chipping sequence itself. The size of watermark is increased n times. Needed bandwidth for the spread signal is n times larger than of the bandwidth of the original signal. The spread signal can improve security if the attacker does not have the spreading code [19]. The concept of DSSS is illustrated in Figure 2.2.

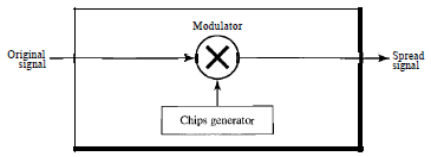

Figure 2.2. DSSS [19].

2.3.2.2. Frequency Hopping Spread Spectrum. The frequency hopping spread spectrum (FHSS) method utilizes M various carrier frequencies which are modulated by the source signal [19]. If there are M different carrier frequencies, for M different time, each of the frequecies is used by order. After the first cycle is finished, the same sequence is repeated. Figure 2.3 illustrates the general diagram for FHSS. A pseudorandom code creator, called pseudorandom noise (PN), generates k-bit pattern for every hopping interval [19].

There are mainly two advantages of this technique [19]. The first advantage is the antijamming affect. If the jammer does not know the hopping period, it can only

attack some little part of the signal. The second advantage is security. Malicious users can not access all data if they do not know the hopping period. They can only access a small part of it.

This method can be applied to watermarking systems. The host signal is

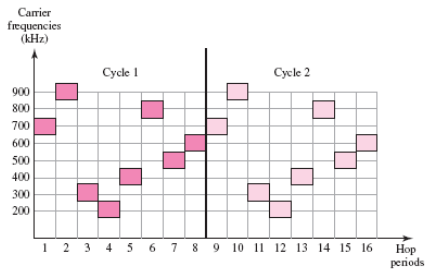

Figure 2.3. FHSS [19].

transformed into frequncy domain by using DCT or DFT. Then, the watermark is embedded to the frequencies which are determined by hopping periods. This process is repeated for time cycles. However, we implemented this technique differently. Every amplitude of the sound can be represented by bytes. Since every byte comprises of eight bits, we used bits as frequencies in FHSS. Length 'n' hopping sequence is created where every element of this sequence is an integer between 1-8. This sequence will tell us which bit to embed the watermark. Here 'n' means embed the watermark to the nth least significant bit.

2.3.3. Transformation Techniques

Here, the background of discrete wavelet transform(DWT) is mentioned.

2.3.3.1. Discrete Wavelet Transform. Time domain is the most common representation used in practical applications. By drawing the time domain signal, time amplitude representation is attained [3]. But, signal representation in time domain is not adequate because it does not give the information about spectrum of frequencies that exists in

the signal.

Frequency domain gives the minutiaes of the frequency elements of the signal. The frequency spectrum of a signal is mainly the frequency elements of this signal [20]. The deficiency of frequency domain is that it loses track of time. In other words, for a given spectrum of frequencies, it does not know to which time interval these frequencies belong.

Time and frequency domains have significant disadvantages. Since only time or only frequency domain does not provide enough information for us, we need to find a domain which carries both information. "Wavelet Transform provides the time-frequency representation of a signal" [3]. There are mainly two different kinds of wavelet transforms. These are continuous wavelet transform (CWT) and discrete wavelet transform (DWT). A complete wavelet system is as follows. First the wavelet transform is applied. Signal goes through some signal processing operations. Then, the inverse wavelet transform is applied. The signal is almost fully recovered.

1-level discrete wavelet transform operation separates the signal as high pass and low pass components. Let us assume we have time domain represented signal S[n]. The DWT process involves passing S[n] through a high pass filter and applying down sampling where we call the result H[n]. And, passing S[n] through a low pass filter and down sampling produce imprecise coefficients that we call the result L[n] [3].

Multi level DWT can be applied in different ways. The first option is to apply one level, then apply second and more on approximate coefficients. Second option is to apply second and more on detailed coefficients. The third option is to apply second or more levels to all transformed signal. Low frequency coefficients are generally selected. The 3-level DWT decomposition is illustrated in Figure 2.4.DWT with multiple levels are applied as in Figure 2.4.

The reconstruction or inverse DWT operation is the exact opposite. The approximate coefficients (L[n]) are up-sampled and passed through a low pass filter and

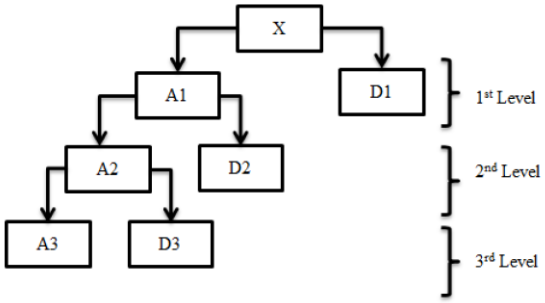

Figure 2.4. Multi-Level DWT [3].

similarly, detailed coefficients (H[n]) are up-sampled and passed through a high pass filter g1[n] [3]. The reconstructed signal is created from the convolution of these two.

2.3.3.2. Haar Wavelet. Although Haar Wavelet is a simple wavelet transformation, it is a useful pedagogical step, and plays an important role in various areas of watermarking, so we use it for an example wavelet transformation. Haar Wavelet Transformation is an orthogonal transformation. 1D Haar Wavelet Transform is used for this purpose [21]. High pass and low pass values are created as the following:

$$H1\,(i) = \frac{S\,(2i) - S(2i+1)}{2} \tag{2.1}$$

$$L1\,(i) = \frac{S\,(2i) + S(2i+1)}{2} \tag{2.2}$$

The inverse transform is the exact opposite of the DWT. The signal is reconstructed as same as the original version.

$$S\left(2i\right) = \ L1\left(i\right) + H1(i) \tag{2.3}$$

$$S\left(2i+1\right) = \ L1\left(i\right) - H1(i) \tag{2.4}$$

3. BIOMETRIC AUTHENTICATION TECHNIQUES

Traditional user authentication or identification systems are interested in something that you possess (like a key, an identification card, etc.) or something you already know (like a password, or a PIN) [22]. With biometrics, this interest has been shifted towards a different approach : something that are part of you (fingerprints or face) or something you make (e.g., handwritten signature or voice) [22]. In areas of users identification needs to be done with a high security, biometric recognition is a rising signal processing area by means of authentication. Traditional authentication systems include something you know. Thus, information like passwords or PINs is subject to being forgotten or lost. In biometric case, this can not happen. Different biometric characteristics exist and each have their own advantages and disadvantages. The choice of which characteristics to use is application dependent.

3.1. Common Biometric Characteristics

Biometric traits can be investigated in two categories:

(i) Physiological

- Face
- Palm
- Fingerprint
- Iris

(ii) Behavioral

- Keystroke
- Signature
- Voice
- Gait

A biometric system is essentially a pattern recognition system which classifies a person by utilizing physical or behavioral traits of a user.

3.2. Biometric Systems

In a biometric system, first we record the users biometric data. From these data, we extract features. This can be expressed as a vector. After that, these features(vectors) are compared with the existing ones in database. The match that has the smallest distance is chosen.

3.3. Different Types of Biometric Techniques

Identification: In an identification biometric system, the system takes a sample and compares with every entry in the database. This is a "one-to-many" (1:N) type of comparison [23]. After comparison by using likelihoods, the system returns the best match or possible matches. The main goal of these type of systems is to identify criminals or terrorists using survaillance.

Verification:Verification system works in such a way that the system obtains one sample and compares with the recorded one [23]. This method is a comparison named one-to-one [23]. The system has two results: match or does not match respectively. List of possible matches, as in identification case, is not returned. Verification is generally used for authentication in computer or digital device access.

3.3.1. Finger Print Recognition

A fingerprint based biometric system is a pattern recognition system that recognizes a person by determining the authenticity of his fingerprint [24]. Every individual person has a unique fingerprint. The fingerprint biometrics is the one of the most studied and commonly used biometric system. The uniqueness of fingerprints has been known and used more than a century. Fingerprint recognition is reinforced with powerful microprocessors and can be used in computers, banking, cars and cellular phone

applications. Wherever we use a key or a password, it can be replaced by our fingerprint [23].

Figure 3.1. A Fingerprint Image.

3.3.2. Face Recognition

The biometric technique which is used to distinguish people from their faces is called face recognition. It analyzes special features in the face such as distance between the eyes, width of the nose, position of cheekbones, jaw line, chin, unique shape, pattern, measurement of the eyes, nose, mouth and other facial features are involved in these systems [23]. Facial expressions such as eye, mouth and lip movement can be combined with the traditional techniques to improve the accuracy of the recognition. Face recognition transforms a digital image from a multimedia source into set of

features that distinguishes people. These features then combined into a single feature which uniquely identifies each person. Face recognition can be done on a whole face or different parts of the face.

Face recognition can be utilized in common applications concerning security. The main application field that differentiates facial recognition from other biometrics is surveillance purposes. It can be used by government authorities to identify criminals, suspected terrorists or missing children. This task is one of the possible usages of the facial recognition system.[23] It can also be used to strengthen the traditional passwords. Additional applications include automated crowd surveillance, access control, face reconstruction, design of human computer interface (HCI), multimedia communication, driver's license and voter registration [23].

To narrow the search in identification, side informations like race, sex , age ,facial expression can be utilized[23]. Identification and verification are two usage areas of face recognition.

Face Recognition can be investigated in dual groups which are Appearance based and Model based, respectively. Appearance based are further classified as Linear which includes Principal Component Analysis (PCA), Independent Component Analysis (ICA) and Linear Discriminate Analysis (LDA) and Non-Linear which includes Kernel PCA (KPCA), ISOMAP, LLE, etc. Model based Face recognition is further classified as 2D and 3D. 2D includes Elastic Bunch Graph and Active Appearance Model [23].

3.3.3. Iris Recognition

High resolution pictures of irises of people's eyes can be used in biometric systems which are called iris recognition systems. The annular area between the pupil and the white sclera in the eye is called the Iris, in other words, it is the colored ring that borders the pupil [23]. It has a rich texture based on interlacing features, which is

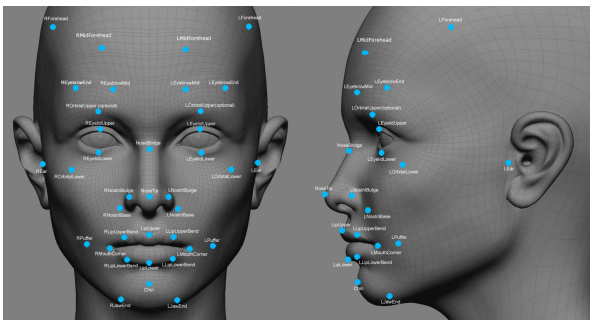

Figure 3.2. Face Recognition Features [25].

called the texture of the iris [23]. This texture is well known to provide a signature that is unique to each individual [26]. Retina scan technology maps the capillary pattern of the retina, a thin nerve on the back of the eye [23]. The difference between retina scan and the iris scan is that retina scanning is interested in the pattern of blood veins whereas thie iris scan catches the iris. Retina scan is used as a high security biometric authentication system in military and government organizations. Through a life time the iris' uniqueness and shape remains unchanged.

After the transformation of the image into the digital form, some mathematical representations are obtained that uniquely identifies a person.

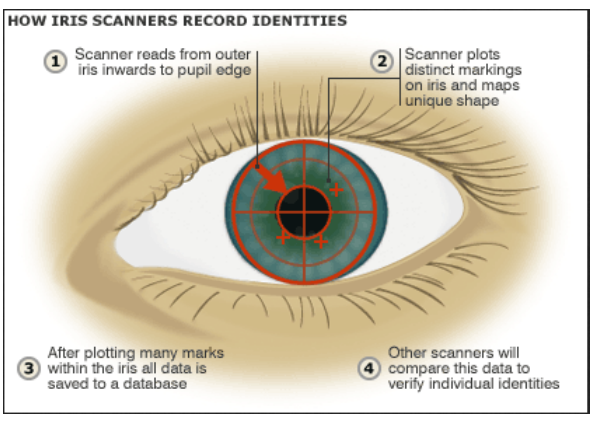

Figure 3.3. Iris recognition procedure [27].

3.3.4. Hand Geometry

Physical qualities of a user's hand and fingers can be used as a biometric authentication system named as Hand Geometry. The bifurcations or branches which are made by the ridges and finger image ridge endings are analyzed [23]. The features extracted from hand to be used in the system are the length, width, thickness and surface area. This biometric technique is easy to use and is widely used due to its cost factor, which is relatively low. It is normally applied in access control and participation [23]. A verification template is constructed by a 3D image of the hand, then this template is saved in the database. For verification purposes, the current template is compared with the existing ones in the database. Depending on the matching score, the user is either accepted or evaluated as the imposter.

3.3.5. Voice Biometrics

The voice of a person can be utilized as a biometric technique for verification or identification purposes. The unique biometric features of a person can be extracted by using a microphone which exists in almost every mobile phone and computer. This biometric technique is generally used in telephone-based applications [23]. The user says a fixed or random phrases into the microphone or telephone handset. Then some distinguishing features are extracted. Voice verification is very low interfering biometric method [28].

There are mainly two types of speaker recognition systems as far as speaker is concerned: open-set and closed-set systems. In closed-set identification the person that will be identified must exist in the training database of voices. On the contrary, open-set identification allows speech input from a person absent in the database [29]. The system specifies the person as unknown speaker.

There are mainly two types of speaker recognition systems as far as content is concerned: text-dependent and text-independent systems. In text-dependent systems, speaker must utter a predefined word or word sequence. On the other hand,

text-independent systems do not have this limitation [29].

In contrast to speaker identification, in speaker verification systems there is only one registered user. The aim is to confirm whether the speaker is the registered speaker or not. Such systems are widely used in security applications to build a multi-level permission system [29].

To review the literature, [30] is selected as a baseline speaker identification system. Reynolds and Rose [30] used MFCC (Mel-frequency Cepstral Coefficients) as a feature vector. To classify speakers, GMM (Gaussian Mixture Models) was used [29]. For model fitting, maximum likelihood approach was used. Furthermore, they measure the effect of number of GMM's on identification accuracy. Many researchers have been developing systems to improve the baseline. There are two main rooms for development. The first one is to change the feature vector. The second one is to change the machine learning algorithm.

Zhou et. al [31] suggest to use LFCC (Linear Frequency Cepstral Coefficients) rather than MFCC, and JFA (Joint Factor Analysis) and PLDA (Probabilistic Linear Discriminant Analysis) are used for model evaluation [29]. The authors deduced that LFCC gives better performance especially for female speakers. Another system for text-independent closed set speaker identification task is proposed by Espy-Wilson et. al. [32]. For the extracted features, they used four formants (F1,F2,F3,F4) periodic and aperiodic energy in the speech, the spectral slope and difference between two harmonics. When compared to MFCC, the results are slightly better.

Selecting the most appropriate algorithm for modeling is vital as feature selection. In Hasan et. al [33], Vector Quantization technique was used. Four vectors per code book was enough to achieve a hundred percent accuracy. Kamruzzaman et. al and Platt have used SVM as machine learning algorithm. Kamruzzaman et. al [34] used MFCCs as features. SVM was utilized for differentiation of speakers. SVM is a promising approach but it is necessary to investigate it more carefully and to test it on

larger database [29].

3.3.6. Keystroke Dynamics

Behaviour and rhythms of the typing characters are used as a biometric authentication system named as Keystroke Dynamics. A unique biometric template is created and recorded by using keystroke dynamics. The most important and most easily measured features that uniquely identifies people are the time of being a key pressed down (dwell time) and the time between pressing two keys (flight time) [17]. The data then saved and used with machine learning algorithms to determine the unique keystroke dynamics of the user. For verification purposes, the current template is compared with the recorded one. The most common area in which keystroke dynamics can be used is the authentication.

Every person has a different typing style and speed. Especially typing speed depends on the individual. For identification purposes, keystroke dynamics makes use of the current typing speed of the user and compare with the saved one. A user who writes 60 words per minutes can be distinguished from the imposter who has half speed. There are several key properties that are used to authenticate the user. The time that the key is pressed down and the time that passes through the switching keys are the most important factors. In addition, some letters are difficult to get for some people. However, these letters depend on the individuals [23]. Right handed people are writing fast if most of the characters are in the right side of the keyboard. This fact is also true for left handed people with left side of the keyboard. Also used finger is another factor that affects the speed.

There are some advantages of Keystroke Dynamics when compared to the other biometric methods. Since Keystroke Dynamics is used along with a PIN or a password to verify identity, it is very resistant to counterfeiting [23]. In addition, in terms of privacy it is better. No vital information of an individual is recorded. It is also more acceptable to users because the user only enters characters to the keyboard. In other words the user does not touch any special device. Keystroke Dynamics is reliable

as it just uses regular keyboards unlike fingerprint scanners which after aged, give increased error rates [23]. The installment cost is almost zero because no additional hardware is needed such as camera or sensor. Keystroke dynamics system works on the keyboard which is used in a personal computer or a smart phone. Like other biometric techniques, Keystroke Dynamics also has some shortcomings since the user's typing might not be always consistent which will result in more False Rejection Rates (FRR) [23]. To prevent high FRR, we can increase the acceptance threshold. However, this causes high FAR rates. The acceptance threshold must be adjusted so that these two rates(FAR and FRR) are minimum.

3.4. Working of Biometrics

All biometric systems are four level procedures [23].

- Capture: Biometric features such as fingerprint, voice etc. are acquired as raw by the biometric system.
- Feature Extraction: The features that uniquely identify a person are extracted from the captured data. These features then transformed into a unique template. The biometric characteristics are represented as digital numbers.For every individual these numbers are stored.
- Comparison: When identification or verification is needed, the current template is compared with the stored ones.
- Match/non-match: If the sample matches with the stored one, the user is identified and access to a system is granted. Otherwise, the user is labeled as imposter and access is not granted.

3.5. Biometric Performance Measures

The performance of a biometric system can be measured using FAR and FRR detailed as follows:

- The false acceptance rate (FAR) is the probability that the system makes a

mistake when matching the input to entry in the database. It measures the ratio of the inputs that the system is granted access but it should not give [35]. In case of similarity scale, if the person is imposter in real, but the matching score is higher than the threshold, then he is treated as genuine that increase the FAR and hence performance also depends upon the selection of threshold value [35].

- The false rejection rate (FRR) is the probability that the system makes a mistake when not matching the input to entry in the database. It measures the ratio of the inputs that the system is not granted access but it should give [36].

FAR and FRR must be as low as possible, but both are antagonists and part of an intricate balancing act [23]. If we make the system hard to access by decreasing the threshold value -reducing the FAR-, imposter can not enter; however, the system may not grant access the true user also by increasing FRR. If we increase threshold so that user gets access every time -decreasing FRR-, then imposters also may access the system-increasing FAR-.

3.6. Application of Biometric Techniques

Biometric systems are new technology. It is commonly used in different application areas. A lot of companies use biometric authentication systems to physically access buildings, doors, areas. Unauthorized access to ATMs, workstations, cellular phones, personal computers has become harder with the usage of biometrics. Telephonic and Internet transactions (e-commerce and e-banking) also use different Biometric techniques [23]. Increasing security threats have forced many countries to start using biometrics for border control and national ID cards [28].

4. PROPOSED TECHNIQUE FOR WATERMARKING

The project of Real-Time Audio Watermarking is explained in this thesis. The purpose of the project is to give a real-time non-detectable and secure audio watermarking system which is implemented on Android based mobile devices. Another aim of the project is to prepare software for mobile devices to embed watermark data into speech. This software enables the user to embed any data type into speech. Any other mobile device which has this software could extract the watermarked information from the conversation. However, third agents should not be available to do the extraction. Moreover, any third agent should not be able to detect that there is watermark in the speech.

The initial aim was to construct the system such that any pre-specified data is automatically embedded into all conversations made on the mobile device in which the application is installed. Therefore, we did not bother to implement a VOIP program but we used one of them, Sipdroid.

4.1. Infrastructure

Since we would need a program which enables voice over IP communication, we used Sipdroid. Moreover, it is an open source program and is widely used. To make a call, we needed a Pbxes account. We took one and made the necessary adjustments to work with it (Trunk, Inbound and Outbound routing and etc.). We downloaded the source code of Sipdroid and investigated it. For more information you can go to the web site http://sipdroid.org/.

In Sipdroid program, RtpStreamReceiver and RtpStreamSender are responsible from sending and receiving the speech. So, in order to embed the watermark we changed these two files. We saved an image to the sdcard of our telephone. After the conversation, the same image is reconstructed at the receiver phone.

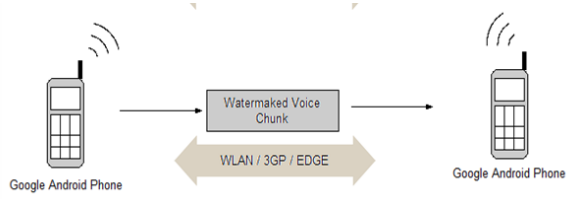

Figure 4.1. Desired Communication Channel.

Two methods are applied. The first one is the Time Domain FHSS Method [37] and the second one is the wavelet domain FHSS and DSSS method [38]. These two methods are explained in what follows.

4.2. Time Domain FHSS Method [37]

4.2.1. Watermark Preprocessing

First of all, the image is saved to the sdcard. From there, the image is retrived pixel by pixel in the RtpStreamSender.java code as follows:

(i) Represent the grey-scale image watermark as a two dimensional MxN matrix. Each element of this matrix represents the pixel value which is between 0-255.

(ii) Convert this matrix into a one dimensional array. The rows will be concatenated after each other. The length of this array is MxN.

(iii) Zeros and ones are needed in order to apply the spread spectrum technique to embed different parts of audio bytes. First we converted the pixel value into an 8 bit binary number. Then, a new array to hold this zeros and ones is created. This new watermark array has length equals to 8 times the old one. This array's elements will be embedded to audio.

(iv) Frequency hopping spread spectrum technique is used. A hopping sequence

is necessary for this purpose. A length five hopping sequence is created where every element of this sequence is an integer between 1-8. This sequence will tell us which bit to embed the watermark. Here 'n' means embed the watermark to the nth least significant bit.

(v) Convert an audio signal into a byte array.

The original byte	Bit of r	Result byte
01100100	r[0]=0	01100<u>1</u>00
11100100	r[1]=1	11100<u>0</u>10

The original byte	Bit of r	Result byte
01100100	r[0]=0	01100<u>0</u>00
11100100	r[1]=1	11100<u>1</u>00

The original byte	Bit of r	Result byte
01100100	r[0]=0	01100<u>0</u>00
11100100	r[1]=1	11100<u>1</u>00

The original byte	Bit of r	Result byte
01100100	r[0]=0	0110<u>0</u>100
11100100	r[1]=1	1110<u>1</u>100

The original byte	Bit of r	Result byte
01100100	r[0]=0	011<u>0</u>0100
11100100	r[1]=1	111<u>1</u>0100

Figure 4.2. Embedding with Hopping Sequence 23345.

4.2.2. Watermark Embedding

In the RtpStreamSender, the Sipdroid program sends the speech over the IP. The buffer which consists of 960 bytes is sent. The embedding is done in these bytes. From the watermark preprocessing, we have an array consisting of zeros and ones. In the buffer, we embed this array into every one byte in four bytes.

(i) Determine the position that the watermark bit will be embedded by using a hopping sequence. This hopping sequence will be repeated. For example, if sixth bit will be embedded and the length of the hopping sequence is five, this bit will be

embedded to the position that the first element of hopping sequence specifies. (6mod5)

(ii) Obtain the embedded watermark bit from the binary array.

(iii) Embed every one byte in four bytes of the audio file.

(iv) Zeros and ones arrays are constructed. They determine which bit we are going to embed. Zeros is for zero embedding, whereas ones for one embedding.

$$zeros[n] = 255 - 2^n \tag{4.1}$$

$$ones[n] = 2^n \tag{4.2}$$

(v) If the element is '0' use bitwise 'AND'. Otherwise use bitwise 'OR' with the audio byte arrays' specified bit. This means we replace the specified bit in the audio byte with the watermark bit. For instance if the embedded bit is '1', embedding will be done as follows:

$$audioBytes[i] = audioBytes[i] \vee ones[n] \tag{4.3}$$

If the bit is zero:

$$audioBytes[i] = audioBytes[i] \wedge zeros[n] \tag{4.4}$$

where audioBytes is the audio array, n is the bit's embedding position.

(vi) Obey the hopping sequence periodically until the last element of the watermark bit array is embedded.

4.2.3. Watermark Extracting

In the RtpStreamReceiver, the Sipdroid program retrieves the speech over the IP. The buffer which consists of 160 bytes is received from the transmitter. The extraction is done in these bytes. From the hopping sequence, we know where the embedded bits are. After we collected all zeros and ones, we convert these into decimal representations of pixels. After that, we created the image at the receivers sdcard.

(i) In the watermarked audio byte array, the embedded bits location is obtained by using the hopping sequence.

(ii) Retrieve the embedded bit by applying bitwise AND operation with the binary number which has all zeros but a one in the embedding position.

$$embeddedBit = (audioBytes[i] \wedge 2^n)/2^n \tag{4.5}$$

(iii) Put the retrieved bits into the bit array.

(iv) Convert these binary representations into decimal representations. This operation will result in the pixel value array (M x N).

(v) Convert the pixel value array into an M x N two - dimensional matrix.

(vi) Represent this resulting 2D matrix as a gray scale image.

4.3. Wavelet Domain FHSS and DSSS Method[38]

4.3.1. Watermark Preprocessing

(i) The grey-scale image watermark is represented as a two dimensional MxN matrix. Each value of this matrix represents the pixel value which is between 0-255. Imread() method at MATLAB is used for this purpose.

(ii) Convert this matrix into a one dimensional array. The rows will be concatenated after each other.The length of this array is MxN.

(iii) Zeros and ones are needed in order to apply the spread spectrum to embed different parts of audio bytes. First we converted the byte value into an 8 bit binary number. Then, a new array to hold this zeros and ones is created. This new watermark array has a length equals to 8 times the old one.

(iv) Direct Sequence Spread Spectrum Technique is applied on the watermarked file. A chipping sequence of length eight is determined. Every '1' is replaced with the negation of the 8 bit chipping sequence whereas each '0' is replaced with the chipping sequence itself. The size of the array is increased 8 times. This new array's elements will be embedded to audio.

(v) Frequency hopping spread spectrum technique is used. Hopping sequence is necessary for this purpose. Length five hopping sequence is created where every element of this sequence is an integer between 1-8. This sequence will tell us which bit to embed the watermark. Here 'n' means embed the watermark to the nth least significant bit.

(vi) Convert an audio signal into a byte array.

(vii) 2 level DWT is applied to the audio byte array. 1D Haar Wavelet Transform

is used for this purpose. High pass and low pass values are created as the following:

$$H1\,(i) = (S\,(2i) - S(2i+1)) * 1/2 \tag{4.6}$$

$$L1\,(i) = (S\,(2i) + S(2i+1)) * 1/2 \tag{4.7}$$

where S is the audio signal. The second level transform is:

$$H2\,(i) = (L1\,(2i) - L1(2i+1)) * 1/2 \tag{4.8}$$

$$L2\,(i) = (L1\,(2i) + L1(2i+1)) * 1/2 \tag{4.9}$$

(viii) Audio byte array is arranged so that L2- H2- H1 is concatenated after each other, respectively.

4.3.2. Watermark Embedding

(i) Determine the position that the watermark bit will be embedded by using hopping sequence. This hopping sequence will be repeated. For example, if sixth bit will be embedded and the length of the hopping sequence is five, this bit will be embedded to the position that the first element of hopping sequence specifies (6mod5).

(ii) Obtain the embedded watermark bit from the binary array.

(iii) Embed every one byte in four bytes of the audio file.

(iv) If the element is '0' use bitwise 'AND'. Otherwise use bitwise 'OR' with the audio byte arrays' specified bit. This means that we replace the specified bit in the audio byte with the watermark bit. For instance, if the embedded bit is '1', embedding

will be done as follows:

$$\text{audioBytes(i)=bitor(audioBytes(i,}2^{n-1}) \tag{4.10}$$

If the bit is zero:

$$\text{audioBytes(i)=bitand(audioBytes(i,255-}2^{n-1}) \tag{4.11}$$

where audioBytes is the audio array, n is the bit's embedding position.

(v) Obey the hopping sequence periodically until the end.

(vi) Apply 2-level Inverse DWT to the audio file.

$$L1\,(2i) = \; L2\,(i) + H2(i) \tag{4.12}$$

$$L1\,(2i+1) = \; L2\,(i) - H2(i) \tag{4.13}$$

$$S^{'}\,(2i) = \; L1\,(i) + H1(i) \tag{4.14}$$

$$S^{'}\,(2i+1) = \; L1\,(i) - H1(i) \tag{4.15}$$

Here $S^{'}$ is the watermarked original signal.

4.3.3. Watermark Extraction

(i) Apply 2 level DWT to the $S^{'}$.

(ii) In the watermarked audio byte array, the embedded bits location is obtained by using hopping sequence.

(iii) Retrieve the embedded bit by applying bitwise AND operation with the binary number which has all zeros but a one in the embedding position.

$$\text{embeddedBit=bitand(audioBytes(i), } 2^{n-1})/\ 2^{n-1} \tag{4.16}$$

(iv) Put the retrieved bits into a bit array.

(v) For every eight element of this retrieved array, XOR with the 8 bit chipping sequence. Then sum up all eight results. If the result is greater than 4, replace these eight elements with a '1'. Otherwise replace with a '0'.

(vi) Convert these binary representations into decimal representations. This operation will result in the pixel value array of size (M x N).

(vii) Convert the pixel value array into an M x N two - dimensional matrix.

(viii) Represent this resulting 2D matrix as a gray scale image. Mat2gray() method in MATLAB is used for this purpose.

5. COLLECTING PERSONAL BIOMETRIC DATA

Biometric classifiers are definite qualities applied to label people. There are two types of biometric identifiers which are physiological and behavioral characteristics. Human physiology is connected to physiological features. classifies individuals by using voice, DNA,face [36]. Behavioral characteristics are connected to the pattern of behavior of an individual, such as typing rhythm, gait, and voice[36]. We implemented one feature for each on Android phone. Before anyone speaks; we collect his biometric data and save them to a text file. After that, we watermark this file into voice.

Figure 5.1. You can either begin to create your biometric data by pressing the button register or try to login with your existing data.

5.1. Registration

Users should be registered with their biometrics information. First of all, the TC ID Number is asked from the user. After that, the password should be entered ten times correctly while keystroke biometrics data is collecting. Then five voice samples in a row are collected. Finally, the user can make the call while sending his biometric data.

5.1.1. Unique TC Identity Number

While creating personal biometric data, the first data we collect is the T.C Identity number. User enters his/her T.C Identity number and this number is used as a username. Because this number is unique, it is a well chosen user name.

Figure 5.2. After entering Identity Number press Confirm and then press Next.

5.1.2. Collecting Keystroke Dynamics Data

Keystroke dynamics is a kind of behavioral biometric technique that emerges with the idea that a unique typing pattern differentiates people. Based on this pattern, people can be identified by their unique typing characteristics [17]. In the 19th century, operators were able to recognize a person over telegraph lines by that person's keying dynamics [17]. So, the same concept can be applied to the commonly used touchscreen cell phone keyboards as a unique signature. Keystroke dynamics has also different names such as keyboard dynamics, keystroke analysis, typing biometrics.

As far as keystroke dynamics is concerned, there are several features that depend on the people; for example, the time passes till they find the keys, how frequent they make mistakes or how long they push the keys. But, the most important and most easily measured features are calculating the time of being a key pressed down (dwell time) and the time between pressing two keys (flight time) [17]. If we can manipulate these timing values, we can apply the keystroke concept to the password authentication and strengthen this type of authentications.

In Keystroke Dynamics, two metrics are used to find the identity of a user: Dwell Time and Flight Time [17]. During the entering of the password, our application records every time a button is pressed or released. From these data we can calculate the flight and dwell time for every character in the password. The user types a verification password that he knows for verification purposes in our application (i.e. account ID and password).

After the user enters the identity number, the user is asked to enter his password. The password is of length 10 digits. If the user enters a password of length less than ten, a warning appears and the user should enter his/her password again. After that, the user enters the password ten times. After the user successfully enters the first attempt, if he enters the password wrong in upcoming attempts, this attempt does not count and the user should enter the password again. We collect the data for every attempt. For every entry we collect data for 10 dwell and 9 flight times. The

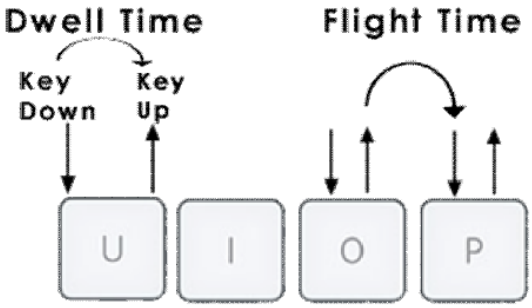

Figure 5.3. Keystroke Dwell Time and Flight Time [17].

collected data consist of a lower threshold, an upper threshold and an error margin. The lower threshold is the user's lowest limit and the upper threshold is his/her upper limit. Users' value should be between these limits to be acceptable. The error margin is used to determine how much these limits can be tolerated.

Estimation of these thresholds is explained as below [17]:

First, the average of the unfiltered values is calculated as follows:

$$D_t = \frac{1}{n} * \sum_{k=0}^{n-1} d_t \tag{5.1}$$

Then the temporary margin of error is determined, which is the highest difference between the average value and unfiltered timing values for both Dwell and Flight times.

$$e_f = (DynamicFlightMarginOfError) \tag{5.2}$$

$$e_d = (DynamicDwellMarginOfError) \tag{5.3}$$

$$e_d = [e_0^d, \ . \quad .,e_{n-1}^d] \tag{5.4}$$

$$e_f = [e_0^f, \ . \quad .,e_{n-1}^f] \tag{5.5}$$

Determine the lower threshold value by subtracting half of the determined temporary margin of error value from the Flight average and Dwell average.

$$Th^d = (DwellThreshold) \tag{5.6}$$

$$Th^f = (FlightThreshold) \tag{5.7}$$

$$Th^d_{Lower} = d_t - \frac{E_{Temp}}{2} \tag{5.8}$$

$$Th^f_{Lower} = f_t - \frac{E_{Temp}}{2} \tag{5.9}$$

Determine the upper threshold value by adding half of the temporary dynamic margin of error to the Flight average and Dwell average

$$Th^d_{Upper} = d_t + \frac{E_{Temp}}{2} \tag{5.10}$$

$$Th^f_{Upper} = f_t + \frac{E_{Temp}}{2} \tag{5.11}$$

Calculate the margin of error of the user, which is the average of all collected dynamic margin of errors for both Dwell and Flight latencies. Divide the estimated default margin of error (125ms) by 4 and add to the sum of Flight and Dwell Dynamic margin of errors.

$$E^{f}_{Dynamic} = \frac{1}{n} * \sum_{k=0}^{n-1} e_f \tag{5.12}$$

$$E^{d}_{Dynamic} = \frac{1}{n} * \sum_{k=0}^{n-1} e_d \tag{5.13}$$

$$E_{User} = E^{d}_{Dynamic} + E^{f}_{Dynamic} + E_{Default} \ /4 \tag{5.14}$$

Figure 5.4. After entering your password press next. After Wrong Attempts a Warning Appears.

5.1.3. Biometric Feature Extraction From Voice

The signal of a speech used in communication has some attributes which are unique to the person's speaking. These properties of the sound signal are extracted and they are used to classify the speaker. This classification method is added to Sipdroid application to make it able to identify the speaker from his sound signal.

Three of these differential sound features are used in this application: zero crossing rate (ZCR), standard deviation (SD) and the average magnitude [16]. Wavelet transform-based method provides thegreatest effectiveness in evaluating these speech characteristics and in this application one level wavelet transform method is used before extracting these attributes from the sound [16].

Figure 5.5. Users record their voice sample five times.

5.1.3.1. Zero Crossing Rate Calculation. The first feature we extracted is the Zero Crossing Rate. Zero crossing rates mean the number of signal polarity variations. In the wavelet transformed signal, zcr value is equal to the rate of sign-changes of a signal, in other words it is the ratio of variation of the signal from postive to negative. The

formula below gives the zero crossing rate:

$$ZCR = \frac{1}{L-1} * \int_{x=1}^{L-1} D\left\{s_{x-1}s_x<0\right\} \tag{5.15}$$

where s and L indicates the signal and the length of it respectively. The D function is 1 if $s_{x-1}s_x$ is less than 0, and 0 otherwise.

5.1.3.2. Standard Deviation Calculation. The second feature we extracted is the Standard Deviation. The formula for the standard deviation is given by:

$$\sigma = \sqrt{MEAN[s_n - \mu]^2} \tag{5.16}$$

where s_n is an amplitude signal and μ is the average amplitude of the signal .

5.1.3.3. Average Magnitude Calculation. The third feature we extracted is the Average Magnitude. The formula for the average magnitude is given by:

$$Average\ Magnitude = \frac{1}{L} * \int_{m=1}^{L} |\mathrm{s}\left(\mathrm{m}\right)| \tag{5.17}$$

where s is the signal and L is the signal's length.

5.1.3.4. Acceptance Interval Calculation. After the registration process, we have 7 voice samples of the user. From each sample, the values of three certain attributes are extracted. These attributes are average magnitude, zero crossing rate and standard deviation as described above.

Now, belonging to this user, we have 7 values of average magnitude, 7 values of zero crossing rate and 7 values of standard deviation. We then specify an acceptance interval representing these 7 values. This interval helps us to decide whether another incoming voice belongs to the same user who registers Sipdroid or not.

After a number of experimental trials, we have decided the interval `[mean-2*var, mean+2*var]` will provide a maximum gain in terms of security and stability. Here, `mean` and `var` are the average and the variance of the 7 values belonging to the related property. When a user tries to login, if the values obtained from his/her voice are in this acceptance interval for all the three properties then he/she is successful to login and otherwise he/she is failed to login.

Figure 5.6. User arrives at calling activity after registration and sends the biometric data.

5.2. Login

Through the registration process, the biometric data is saved to a text file. When a user tries to login the system the entered information is compared with the previous entered ones. Text file contains information for each line, in order to parse this file easily. The currently entered data will be checked whether the data are in the interval that is determined during registration.

5.2.1. TC Identity Number Check

The entered number is checked with the registered TC Identity Number. If the number is not matched a warning message appears and the same activity is called again. If there is a match, then the user can pass the keystroke dynamics authentication.

5.2.2. Keystroke Dynamics Authentication

The password is asked from the user. A two layer security is implemented by requiring the password. If the currently entered password is not equal to the registered one, then an error message appears. User can not continue to voice authentication. This is the first layer of security.

As a second layer, if the password matches and the keystroke dynamics does not match; in other words, the keystroke behavior is different, then another warning appears: “Password matches but pattern differs.” This is the pattern difference case.

5.2.3. Voice Authentication

The user is asked give his voice sample. After that we collect the three features from the voice and check whether these values are in the interval we have calculated at the registration. If the voice’s extracted features are between the calculated intervals, the user matches and goes to the call activity. Otherwise, “Different Voice” warning will pop up.

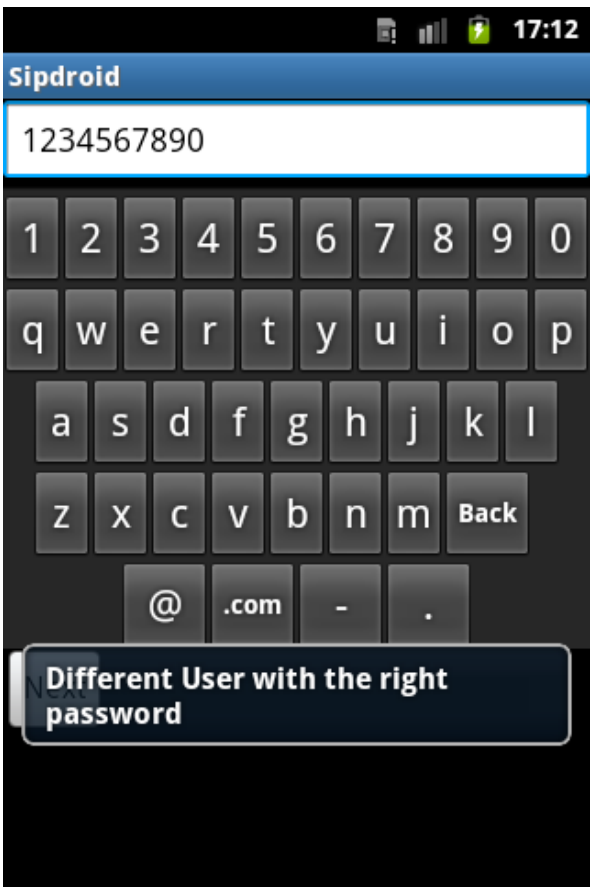

Figure 5.7. The message indicating that password is right but behaviour is different.

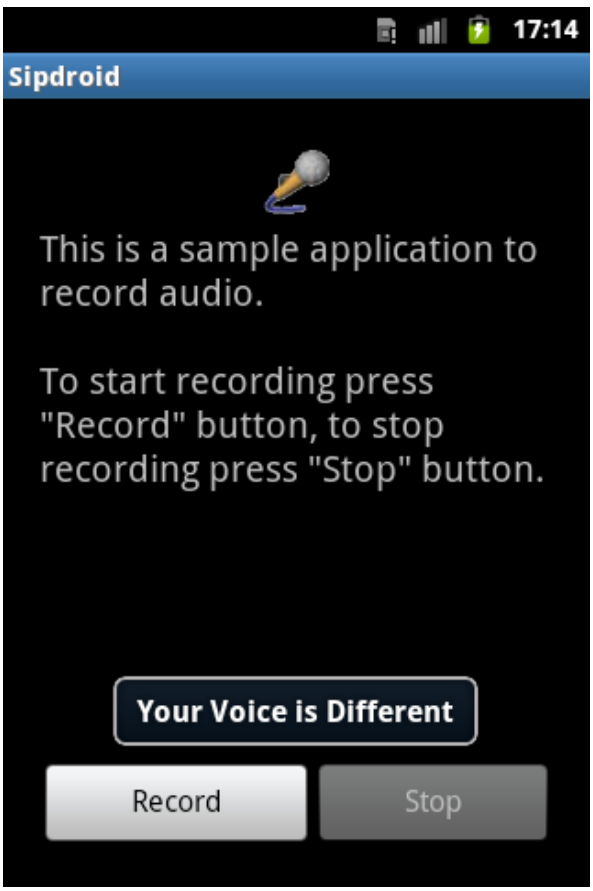

Figure 5.8. The message indicating that voice of user is different.

6. EXPERIMENTS AND RESULTS

Our system is comprised of two parts and evaluation techniques are different for each part. To measure the performance of a watermarking method, there are different metrics such as robustness, imperceptibility and data rate. These metrics are application dependent. For our application, we chose inaudibility and robustness as metrics because we embed image into telephone conversation and these watermarks should not be perceived. In addition, in telecommunication systems, voice is subject to vary common signal processing operations and we want to extract our watermark in these conditions, robustness is also important in our application.

For the second part, biometric authentication, there are two important metrics. The first one is FAR (False Acceptance Rate) and the second one is FRR (False Rejection Rate). We need to come up with thresholds that minimize both of these two metrics. FAR means that we accept some data that does belong to imposter. FRR means we reject some data of the original owner.

6.1. Watermarking System Experiments

To measure the performance of the recommended audio watermarking scheme, a 90 second mono speech file was used. Sampling rate is 88000 sample per second. 8 bits are used to represent every sample.

Embedded image into audio file is 128x128 grayscale Lena image. The length of this image is constant. The audio clip size was adjusted so that watermarking has finished after the processing image bytes and audio has finished simultaneously. The metrics that are applied for measuring the performance of the recommended audio watermarking scheme are inaudibility and robustness.

6.1.1. Inaudibility

Inaudibility (imperceptibility) can be defined as the intuitive difference of the watermark embedded host data from the original audio signal. The goal of inaudibility is that the quality of host signal is not corrupted much, so that the listener can not perceive the difference [9]. Signal-to-Noise Ratio (SNR) is employed for an objective measurement of imperceptibility. Signal to Noise Ratio (SNR) is a difference metric that is used to calculate the similarity between the original audio signal and the watermarked audio signal. The SNR computation is carried out according to the below equation:

$$SNR\,(dB) = 10\log \frac{\sum_n {I_n}^2}{\sum_n (E_n - I_n)^2} \tag{6.1}$$

where I_n is the original audio signal, and E_n corresponds to the watermarked audio signal [10].

Audio SNR results of proposed methods and Audio Watermarking with Wavelet are compared in Table 6.1. As it can be observed from the table, SNR of host audio file is the highest at Time Domain FHSS technique. This means that the effect of changes in time domain is more inaudible. Also, FHSS and Wavelet Domain in Wavelet technique is the most audible. This is because, sacrifice has been made from audibility to spread the bits more on wavelet domain.

Table 6.1. SNR of Watermarked Audio.

	SNR of Watermarked Audio
Audio Watermarking Using Wavelets[11]	56.26
Time Domain FHSS[37]	65.46
Wavelet Domain FHSS and DSSS[38]	54.32

6.1.2. Robustness

Widespread signal processing attacks such as linear filtering, lossy compression, shearing and AWGN are implemented on watermarked audio digital signal. Even though effects of these attacks might not be perceived from the quality of the host signal, embedded watermarking image is more impaired by these attacks.

The similarity between original and extracted watermark from the audio that signal processing operations are applied, is found out by using correlation factor ρ, which can be calculated as follows:[10]

$$\rho(\ w,\ e) = \frac{\sum_n (w_n * e_n)}{\sqrt{\sum_n {w_n}^2} * \ \sqrt{\sum_n {e_n}^2}} \tag{6.2}$$

where e's are extracted image pixels and w's are watermarked image pixels.

6.1.2.1. Compression Attack. Compression attack is applied on both proposed algorithms and the Pure DWT Algorithm. In Figure 6.1 vertical axis shows the correlation factor and horizontal axis shows compression ratio.

Wavelet Domain FHSS and DSSS method's correlation values are the highest for all the compression ratios. In addition, the decreasing rate of wavelet domain methods are accelerating as the compression ratio increases. On the contrary, the slope of Time Domain FHSS method is decreasing because on time domain FHSS algorithm works better at high noisy environments than its competitors. This shows that at low noises wavelet domain is more adventageous. However, when we increase the noise, pure wavelet method becomes unusable. The best combination is the FHSS wavelet combination.

6.1.2.2. AWGN Attack. An Additive White Gaussian attack is applied on both proposed algorithms and Pure DWT Algorithm. In Figure 6.2, the vertical axis shows Normalized Correlation, whereas horizontal axis shows the SNR value of the AWGN.

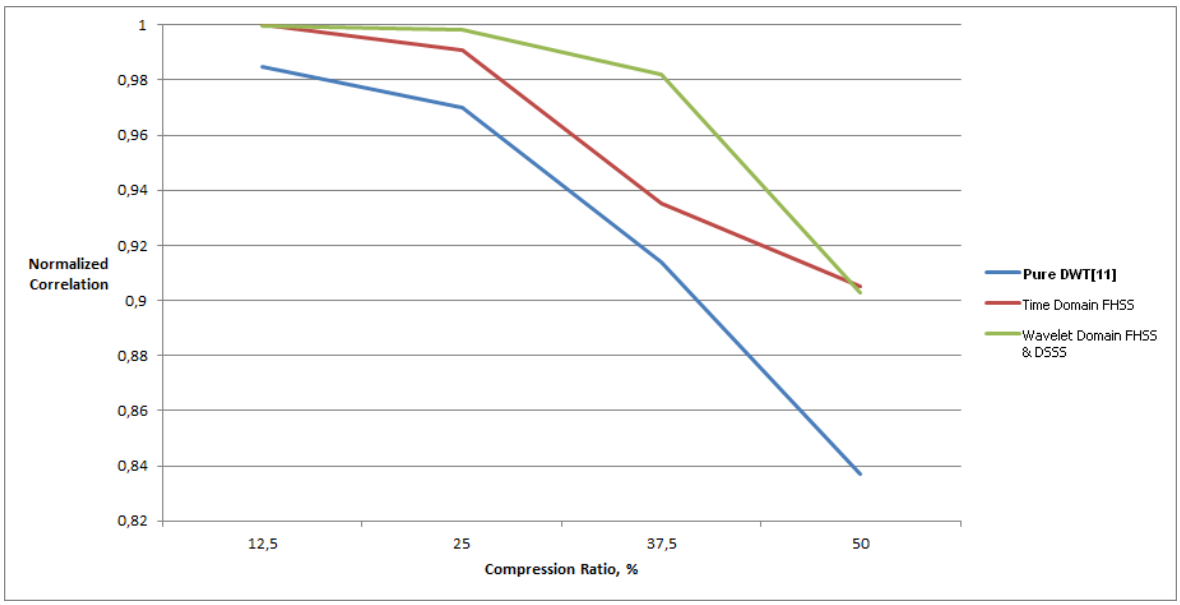

Figure 6.1. Correlation Value versus Compression Ratio.

Correlation values of Wavelet Domain FHSS and DSSS are the highest while we increase the noise. The correlation factor of wavelet domain methods starts higher than Time Domain FHSS method (near 1) when AWGN is 0dB. When the AWGN SNR approaches -10dB, correlation factors of the pure DWT and Time Domain FHSS are starting to become closer. They become equal at -10dB. After that, Time Domain FHSS method's correlation factors are higher than the pure DWT's.

It can be observed from Fig 6.2 that Time Domain FHSS method's correlation factors are smaller at low noises. The decrease in correlation factors as noise increases is approximately linear. On the contrary, the DWT method's correlation factors start with higher values at low noise. The decrease in correlation factors as noise increases is approximately exponential. It can be concluded that the wavelet domain causes an exponential decrease. In addition, FHSS and DSSS embedding increase robustness significantly when compared to pure DWT embedding method.

6.1.2.3. Shearing and Low Pass Filter Attacks. Shearing and Butterworth attacks are also applied. Since the FHSS and DSSS spreads the information 8 times more than the

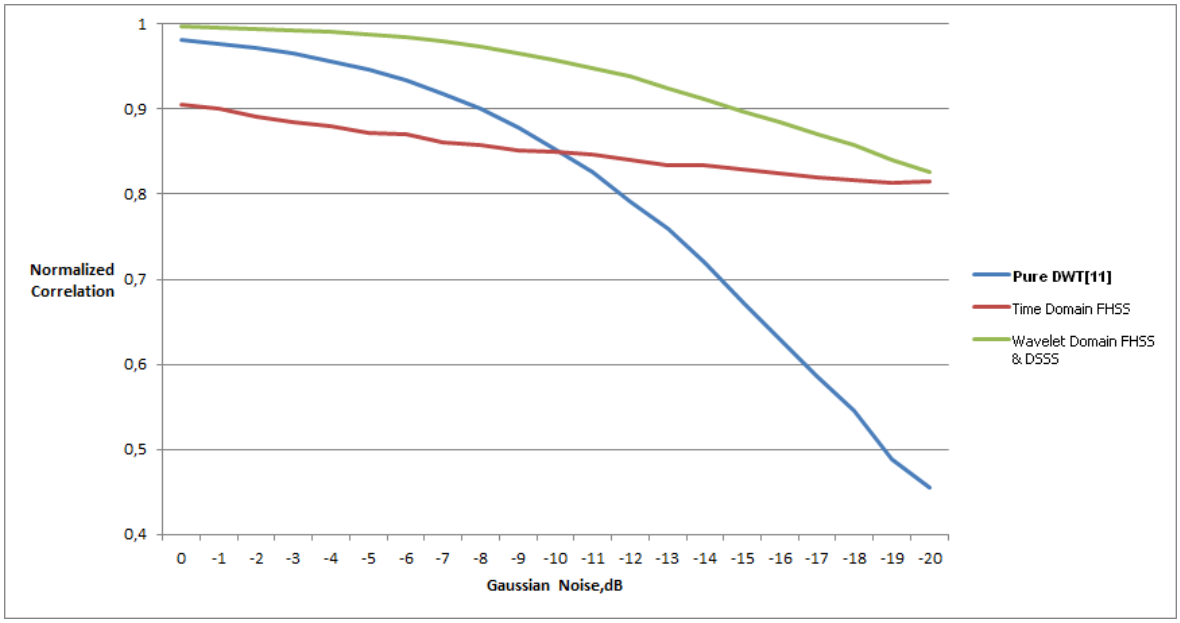

Figure 6.2. Normalized Correlation versus Gaussian Noise (dB).

other embedding technique, the shearing (cut) time was hold 8 times more than that of the DWT technique. The results are shown in Fig. 6.3. The Pure DWT Method is not shown in the Figure, because the value is 0.2542 and is out of the proposed methods' ranges.

A second order Butterworth low pass filter with a cutoff frequency of 0.5 was used. The correlation factors are given in Table 6.2. From the table, it can be observed that for low pass filter attack, Wavelet Domain FHSS and DSSS method outperforms its competitors. Pure DWT technique is out of range with respect to the proposed methods with regarding robustness.

6.2. Biometric Authentication System Experiments

6.2.1. FAR and FRR Rates

We collected 15 voice samples from each of 4 people. And according to the acceptance interval, we calculated their false accept and false reject rates. The values of

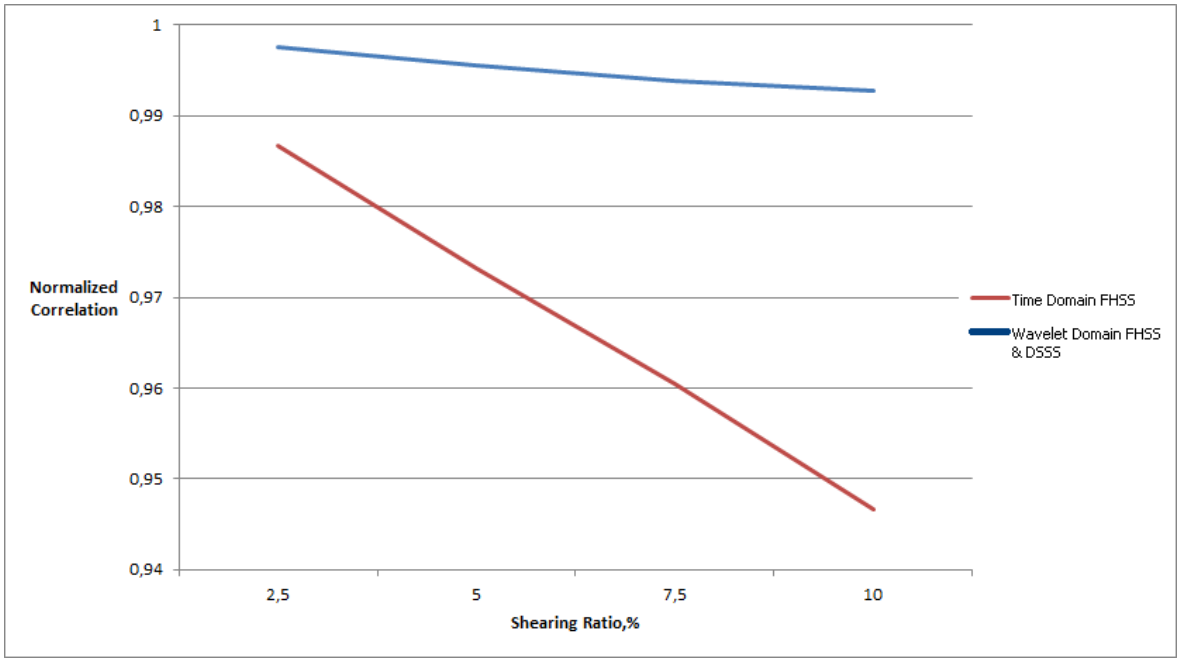

Figure 6.3. Correlation Values against Shearing Attack Ratio (%).

Table 6.2. Correlation Values When LPF and Shearing Applied.

	Butterworth LPF
Pure DWT	0.1203
Time Domain FHSS [37]	0.7642
Wavelet Domain FHSS and DSSS [38]	0.7731

each trial for different users are given at Figure 6.4:

Average Magnitude	Zero Crossing Rate	Std. Dev.	Average Magnitude	Zero Crossing Rate	Std. Dev.	Average Magnitude	Zero Crossing Rate	Std. Dev.	Average Magnitude	Zero Crossing Rate	Std. Dev.
2321	0,06	65,284	1530	0,066	51,913	2644	0,115	67,164	3638	0,047	78,949
2255	0,084	63,095	1503	0,074	55,009	1536	0,11699	49,376	3801	0,053	83,132
2071	0,082	57,862	1603	0,073	55,399	2038	0,119	58,455	3763	0,058	84,682
1765	0,084	59,515	1620	0,068	55,812	1696	0,1049	52,355	3964	0,062	85,223
1874	0,086	59,624	1390	0,081	52,972	1874	0,11824	55,055	3937	0,059	85,563
1894	0,086	57,879	1797	0,068	56,241	1739	0,1284	52,593	3494	0,064	81,08
1532	0,08	52,182	2261	0,068	65,077	1921	0,10895	54,157	3563	0,053	81
1776	0,084	57,018	2393	0,073	65,651	1698	0,1142	51,807	4793	0,058	92,293
1676	0,083	55,281	2654	0,075	69,771	1696	0,12323	52,536	3434	0,054	77,846
1210	0,076	44,215	1930	0,079	57,359	2001	0,1168	56,454	4492	0,08	92,212
1334	0,087	46,819	1962	0,068	55,803	2480	0,1132	61,335	2249	0,058	68,935
1331	0,076	47,666	1796	0,06	54,982	2427	0,1215	62,746	3036	0,08	74,424
1330	0,073	49	1352	0,079	48,959	2247	0,1148	59,498	3457	0,055	78,205
1262	0,081	48,332	1684	0,079	54,745	2245	0,1208	58,634	3559	0,043	82,371
1308	0,08	47,223	1805	0,071	57,096	1592	0,1211	49,759	3311	0,058	77,91
User 1			User 2			User 3			User 4		

Figure 6.4. Extracted Features From Four Users.

False Acceptance Rate (FAR) is the probability that the system makes a mistake when matching the input to entry in the database [35]. It measures the ratio of the inputs that the system is granted access but it should not give [35].

False rejection rate (FRR) is the probability that the system makes a mistake when not matching the input to entry in the database [36]. It measures the ratio of the inputs that the system is not granted access but it should give [36]. And the results are given Table 6.3 and 6.4:

Table 6.3. False reject rate (FRR).

	Number of rejecting himself/herself (out of 15 trials)
User 1	1
User 2	2
User 3	2
User 4	3

From Table 6.3 we found that the average FRR is 2/15 = 0.13.

Table 6.4. False Acceptance Rate (FAR).

	Number of trials with accepting other 3 users ***(out of 15 trials)***			Average Accepting Number
User 1	12	0	0	4
User 2	7	0	0	2,33
User 3	0	0	0	0
User 4	0	1	0	0,33

From Table 6.4 we found that the average FAR is 1.667/15 = 0.11

7. CONCLUSION

In this study, a real time spread spectrum audio watermarking algorithm is implemented. Users should have android smartphones to be able to use the application. As a VoIP application, an open source Sipdroid application is chosen and watermarking algorithm is added to this program. User must copy an image to the sdcard of his/her smart phones. This image is then transferred through IP channel. After that the image is watermarked into voice.

Furthermore, by adding user interface into sipdroid, a biometric authentication system is constructed. Three layers of authentication is needed. The first layer is the T.C Identity number. Since this number is unique, it is used as a user name. In the second layer password and keystroke dynamics authentication are used. Then, the user is asked to enter a password. In this layer, the user should know the password. Also, the behaviour of entering the password should match. The last layer is the speaker identification layer.

To evaluate the spread spectrum based audio watermarking technique on time domain, robustness and inaudibility metrics are used. The DWT based audio watermarking method is also implemented and compared with proposed algorithms. By examining the results, it can be said that inaudibility and robustness performance targets can be obtained. As far as inaudibility is concerned, Time Domain FHSS method gives better results. When we evaluate the robustness performance against signal distortions, Wavelet Domain FHSS and DSSS algorithm definitely outperforms its competitors. Another advantage of the proposed methods to Pure DWT [11] is blindness. These methods do not require the original audio file to extract watermark. Furthermore, it was observed at compression and AWGN attacks that, the difference between the Wavelet Domain FHSS and DSSS algorithm and Pure DWT [11] becomes more clear when attack's power is increased. Our algorithms work much better than its competitors at noisy environments.

To evaluate the performance of the biometric authentication system, FAR and FRR system was used. We calculated these rates for a speaker verification system. Rates were calculated for four people. Threshold was chosen to minimize both rates. Both FAR and FRR had plausible values.

□

REFERENCES

1. Cox, I., M. Miller and J. Bloom, *Digital Watermarking*, Academic Pressing, San Diego,CA,USA, 2002.

2. Zhang, P., S. Xu and H. Yang, "Robust and Transparent Audio Watermarking Based on Improved Spread Spectrum and Psychoacoustic Masking", *Information Science and Technology (ICIST),International Conference on*, pp. 640–643, 2012.

3. Ravula, R., *Audio Watermarking Using Transformation Techniques*, M.S. Thesis, Louisiana State University, 2010.

4. Wang, X., P. Niu and H. Yang, "A Robust Digital Audio Watermarking Based on Statistics Characteristics", *Elsevier Ltd., Pattern Recognition*, Vol. 42, No. 11, pp. 3057–3064, 2009.

5. Kumar, M. N., *Watermarking Using Decimal Sequences*, M.S. Thesis, Louisiana State University, 2004.

6. Baranwal, N. and K. Datta, "Comparative Study of Spread Spectrum Based Audio Watermarking Techniques", *Recent Trends in Information Technology (ICRTIT),International Conference on*, pp. 896–900, 3-5 June 2011.

7. *Digital Watermarking Alliance*, 2012, `http://www.digitalwatermarkingalliance.org/app_broadcast.asp`, accessed at October 2013.

8. Cvejic, N., *Algorithms for Audio Watermarking and Steganography*, Ph.D. Thesis, University of Oulu, 2004.

9. Baoyuan, C., H. Yanli, L. Ruigang and H. Gang, "The Audio Watermarking System Based on Wavelet Transform Algorithm", *Cross Strait Quad-Regional Radio Science and Wireless Technology Conference (CSQRWC)*, Vol. 2, pp. 1274–1277,

26-30 July 2011.

10. Yiping, M. and H. Jiqin, "Audio Watermark in Dct Domain Strategy and Algorithm", *Chinese Journal of Electronics*, Vol. 1, pp. 1260–1264, 2003.

11. Al-Haj, A., L. Bata and A. Mohammad, "Audio Watermarking Using Wavelets", *Networked Digital Technologies (NDT)*, pp. 398–403, 28-31 July 2009.

12. Bartmann, D., I. Bakdi and M. Achatz, "On the Design of an Authentication System Based on Keystroke Dynamics Using a Predefined Input Text", *International Journal of Information Security and Privacy*, Vol. 1, No. 2, pp. 1–12, 2007.

13. Kirovski, D. and H. Malvar, "Robust Spread-Spectrum Audio Watermarking", *Proceedings of IEEE Digital Signal Processing Workshop*, p. 1345–1348, April 2001.

14. Maric, S., I. Seskar and E. L. Titlebaum, "On Cross Ambiguity Properties of Welch-Costas Arrays", *Proceedings of IEEE Transactions on Aerospace and Electronic Systems*, Vol. 30, p. 1063–1071, 1994.

15. Vawter, J., *Audio Watermarking*, Ph.D. Thesis, University of Victoria, 2007.

16. Sheikhan, M., M. Safdarkhani and D. Gharavian, "Presenting and Classification Based on Three Basic Speech Properties, Using Haar Wavelet Analyzing", *Signal Processing Systems (ICSPS),International Conference on*, Vol. 3, pp. 189–191, 5-7 July 2010.

17. Bassia, P., I. Pitas and N. Nikolaidis, "Robust Audio Watermarking in The Time Domain", *IEEE Transactions on Multimedia*, Vol. 3, pp. 232–241, 2001.

18. Arnold, M., M. Schmucker and S. D. Wolthusen, *Techniques and Applications of Digital Watermarking and Content Protection*, Artech House, INC, Boston, London,England, 2003.

19. Forouzan, B., *Data Communications and Networking*, McGraw-Hill, New

York,NY,USA, 2006.

20. Polikar, R., *Home page - Dr. Robi Polikar*, 2001, http://users.rowan.edu/ polikar/WAVELETS/WTtutorial.html, accessed at October 2013.

21. Yan, D. and R. Wang, "Data Hiding for Audio Based on Piecewise Linear Haar Transform", *Image and Signal Processing (CISP), Conference on*, Vol. 5, pp. 688–691, 27-30 May 2008.

22. Ortega-Garcia, J., J. Bigun, D. Reynolds and J. Gonzalez-Rodriguez, "Authentication Gets Personal with Biometrics", *Signal Processing Magazine, IEEE*, Vol. 21, No. 2, pp. 50–62, 2004.

23. Rites, D., *Keystroke Dynamics for Mobile Devices- Data Collection*, M.S. Thesis, SanDiego State University, 2011.

24. Maio, D. and D. Maltoni, "Direct Gray-Scale Minutiae Detection in Fingerprints", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, pp. 27–39, 1997.

25. *Facial Recognition Data Points*, 2012, http://endthelie.com/wp-content/uploads/2012/08/facial-recognition-data-points.jpg, accessed at October 2013.

26. Daugman, J., "Iris Recognition", *American Scientist*, Vol. 89, pp. 326–333, 2001.

27. *Biometric Technology*, 2012, http://news.bbc.co.uk/2/shared/spl/hi/guides/456900/456993/html/nn3page1.stm, accessed at October 2013.

28. Laha, J., *Biometric Techniques -Enhancing Security Standards In High Performance Enterprise*, 2008, http://ezinearticles.com/?Biometric-Techniques--Enhancing-Security

-Standards-In-High-Performance-Enterpriseid=1224599, accessed at October 2013.

29. Sidorov, M., A. Schmitt, S. Zablotskiy and W. Minker, "Survey of Automated Speaker Identification Methods", *Intelligent Environments (IE)*, pp. 236–239, 16-17 July 2013.

30. Reynolds, D. and R. Rose, "Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models", *IEEE Trans. Speech Audio Process.*, Vol. 3, No. 1, pp. 72–83, 1995.

31. Zhou, X., D. Garcia-Romero, R. Duraiswami, C. Esply-Wilson and S. Shamma, "Linear versus Mel Frequency Cepstral Coefficients for Speaker Recognition", *IEEE Automatic Speech Recognition and Understanding Workshop*, 2011.

32. Espy-Wilson, C. Y., S. Manocha and S. Vishnubhotla, "A New Set of Features for Text-Independent Speaker Identification", *ICSLP Interspeech*, 2006.

33. Hasan, M. R., M. Jamil, M. G. Rabbani and M. S. Rahman, "Speaker Identification Using Mel Frequency Cepstral Coefficients", *Electrical Computer Engineering (ICECE),International Conference on*, 2004.

34. Kamruzzaman, S. M., A. N. M. R. Karim, M. S. Islam and M. E. Haque, "Speaker Identification Using MFCC-Domain Support Vector Machine", *International Journal of Electrical and Power Engineering*, Vol. 1, No. 3, pp. 274–278, 2007.

35. Sahoo, S. K., M. Prasanna and T. Choubisa, "Multimodal Biometric Person Authentication : A Review", *IETE Technical Review*, Vol. 29, No. 1, pp. 54–75, 2012.

36. *Biometrics*, 2010, http://en.wikipedia.org/wiki/Biometrics, accessed at October 2013.

37. Can, Y. S., F. Alagöz and M. E. Burus, "A Novel Spread Spectrum Digital Audio

Watermarking Technique", *Journal of Advances in Computer Networks*, Vol. 2, No. 1, pp. 6–9, 2014.

38. Can, Y. S. and F. Alagöz, "Robust Frequency Hopping and Direct Sequence Spread Spectrum Audio Watermarking Technique on Wavelet Domain", *Electronics Computer and Computation (ICECCO),International Conference on*, 7-9 November 2013.